LabPhon 14 - The 14th Conference on Laboratory Phonology

Tokyo, July 25-27, 2014

What's New

2014.08.11 The slides of Prof. Mazuka's talk are now available.

2014.08.08 The slides of Prof. Bob Ladd's talk at the pre-conference colloquium are available now.

2014.08.04 The slides of Prof. Den's talk are now available.

2014.07.29 LabPhon14 has ended. We would like to thank all participants for coming to Tokyo and for making the conference a great success!

Welcome to LabPhon14!

Welcome to the website of LabPhon 14, The 14th Conference on Laboratory Phonology that will take place from 25 to 27 July 2014 at the National Institute for Japanese Linguistics (NINJAL) in Tokyo, Japan.

Conference Themes

The general theme of the conference is **Laboratory Phonology beyond the Laboratory: Quantitative Analyses of Speech Produced outside the Phonetics Laboratory**. Special sessions will be held on:

- fieldwork-based studies of endangered languages/ dialects
- corpus-based approaches to spontaneous speech
- acquisition of L1 phonology/prosody

News and Updates

The conference poster is available here.

LabPhon 14 has ended

LabPhon14 has ended. We would like to thank all participants for coming to Tokyo and for making the conference a great success!

Sitemap

10-2 Midori-cho, Tachikawa City, Tokyo, Japan 190-8561

LabPhon 14 - The 14th Conference on Laboratory Phonology

Tokyo, July 25-27, 2014

2013

2013.11.19
Abstract template and submission form are now available!

2013.06.12
The Call for Papers is now open!

2013.06.12
LabPhon 14 - Website is now available!

Updated Information
2013

Back to home Back Top

Sitemap

10-2 Midori-cho, Tachikawa City, Tokyo, Japan 190-8561

LabPhon 14 - The 14th Conference on Laboratory Phonology

Tokyo, July 25-27, 2014

Guest Speakers

We are pleased to announce the following guest speakers

- **Yasuharu Den** (Chiba University/NINJAL) [abstract]

  "Some cognitive factors behind vowel lengthening in spontaneous Japanese: A corpus-based study" [slides]

  Commentator: Shu-Chuan Tseng (Academia Sinica)

- **Carlos Gussenhoven** (Radboud University Nijmegen) [abstract]

  "On establishing the existence of word stress" [slides]

  Commentator: Aditi Lahiri (University of Oxford)

- **Mark Hasegawa-Johnson** (University of Illinois at Urbana-Champaign) [abstract]

  "Labeling in the wild: Crowdsourcing versus categorical perception" [slides]

  Commentator: Natasha Warner (University of Arizona)

- **Reiko Mazuka** (Duke University/RIKEN Brain Science Institute) [abstract]

  "Infant-directed speech as a window into the dynamic nature of phonology" [slides]

  Commentator: Catherine Best (University of Western Syndey)

LabPhon 14
- Guest Speakers
- Important Dates
- Call for Papers
- Program
- Registration
- Information for Presenters
- Information for Participants
- Accommodation and Travel
- Venue & Maps
- Committees

Back to home   Back   Top

Sitemap

10-2 Midori-cho, Tachikawa City, Tokyo, Japan 190-8561

LabPhon 14 - The 14th Conference on Laboratory Phonology

Tokyo, July 25-27, 2014

Pre-Conference Colloquium

Date

- **Thursday July 24, 2014, 14:30-17:30**
- **Registration for the main conference will also be possible on this day from 13:30-17:30.**

Guest Speakers

- **Bob Ladd (University of Edinburgh)**

"Leaky phonology and the design of language" [slides]

Abstract:
Recent work on sign language - notably the work of Sandler and colleagues on Al-Sayyid Bedouin Sign Language - has shown that phonology emerges in a new language. There does not appear to be a clear transition point from not having phonology to having it. Rather, the phonological organisation of signs becomes gradually more systematic, and even mature sign languages continue to show a variety of features that are difficult to analyse in terms of a smallish finite inventory of phonological elements. In this paper I suggest that the same is true of spoken language phonology as well, and that spoken languages exhibit phonological phenomena analogous to the hard-to-analyse features in sign phonologies. Sapir famously said that "all grammars leak". We are used to thinking that phonology does not leak, and that all phonetic aspects of an utterance can be analysed in terms of a language's finite inventory of phonological elements, independently of grammatical structure. "Leaky" aspects of spoken language phonology are usually idealised out of consideration in various ways (e.g. as paralinguistic or expressive, as dialect mixture or unassimilated borrowing, as historical change in progress), or simply ignored. I discuss several such phenomena, and suggest that they are comparable to what we find in sign languages. The existence of such phenomena argues against a conception of language design in which phonology is autonomous and unconnected to grammar and meaning (the design feature called "duality of patterning" by Hockett and "double articulation" by Martinet). Rather, phonology is in the first instance a property of the internal structure of signs (or words or morphemes), and it is not always possible or desirable to distinguish phonological from grammatical aspects of this internal structure. If phonology is viewed in this light, it is not surprising that it can "leak" like grammar.

- **Anne Cutler (University of Western Sydney)**

"DADDY, EDDY, NINNY, NANNY and BALDEY: Big Data for speech perception" [slides]

Abstract:
In its concluding decade, the Comprehension Group at the Max Planck Institute for Psycholinguistics created several very large sets of speech perception data, all of which are publicly available for the use of any interested parties (such as, for instance, the laboratory phonology community). The Database of Dutch Diphones (DADDY; Smits, Warner, McQueen & Cutler, JASA, 2003) comprises more than a half a million identification responses, by 18 listeners presented with gated fragments of diphones representing every phonetic segment of Dutch in every possible left and right phonetic

context. Even larger is DADDY's English counterpart, the English Diphones Database (EDDY; Warner, McQueen & Cutler, JASA, 2014) which used the same technique with American English input and 22 listeners. The Noise-masked Identifications by Native and Non-Native listeners data set (NINNY; Cutler, Weber, Smits & Cooper, JASA, 2004) contains identification responses by American English (native) and Dutch (non-native) listeners to American English vowels and consonants presented under three levels of noise masking. The other two data sets do not concern phonemic identifications. One is a large corpus of input to 11-month-old infants learning Dutch, with speech from multiple caregivers to the infant and to other adults (NANNY; Johnson, Lahey, Ernestus & Cutler, JASA, 2013). Finally, the BALDEY data set (Ernestus & Cutler, submitted 2014) contains lexical decisions to 2780 spoken words of Dutch and a like number of spoken nonwords. Indicative analyses will illustrate the possibilities of these multifaceted data sets.

Afterwards

17:30- : Informal wine reception

LabPhon 14 - The 14th Conference on Laboratory Phonology

Tokyo, July 25-27, 2014

Guest Speakers

We are pleased to announce the following guest speakers

- **Yasuharu Den** (Chiba University/NINJAL) [abstract]

  "Some cognitive factors behind vowel lengthening in spontaneous Japanese: A corpus-based study" [slides]

  Commentator: Shu-Chuan Tseng (Academia Sinica)

- **Carlos Gussenhoven** (Radboud University Nijmegen) [abstract]

  "On establishing the existence of word stress" [slides]

  Commentator: Aditi Lahiri (University of Oxford)

- **Mark Hasegawa-Johnson** (University of Illinois at Urbana-Champaign) [abstract]

  "Labeling in the wild: Crowdsourcing versus categorical perception" [slides]

  Commentator: Natasha Warner (University of Arizona)

- **Reiko Mazuka** (Duke University/RIKEN Brain Science Institute) [abstract]

  "Infant-directed speech as a window into the dynamic nature of phonology" [slides]

  Commentator: Catherine Best (University of Western Syndey)

LabPhon 14
- Guest Speakers
- Important Dates
- Call for Papers
- Program
- Registration
- Information for Presenters
- Information for Participants
- Accommodation and Travel
- Venue & Maps
- Committees

Back to home Back Top

Sitemap

10-2 Midori-cho, Tachikawa City, Tokyo, Japan 190-8561

LabPhon 14 - The 14th Conference on Laboratory Phonology

Tokyo, July 25-27, 2014

LabPhon 14

10-2 Midori-cho, Tachikawa City, Tokyo, Japan 190-8561

LabPhon 14

The 14th Conference on Laboratory Phonology

General Theme: *Laboratory Phonology beyond the Laboratory: Quantitative Analyses of Speech Produced outside the Phonetics Laboratory*

Specific Topics:
- Fieldwork-based studies of endangered languages/dialects
- Corpus-based approaches to spontaneous speech
- L1 prosody/phonology acquisition

Guest Speakers:

Yasuharu Den (Chiba University/NINJAL)

Carlos Gussenhoven (Radboud University Nijmegen)

Mark Hasegawa-Johnson (University of Illinois at Urbana-Champaign)

Reiko Mazuka (Duke University/RIKEN Brain Science Institute)

Pre-conference colloquium: July 24, 2014

Main conference: July 25-27, 2014

Satellite workshop: July 28, 2014

Place: National Institute for Japanese Language and Linguistics (NINJAL), Tachikawa, Tokyo

http://www.ninjal.ac.jp/labphon14/

国立国語研究所

LabPhon 14 - The 14th Conference on Laboratory Phonology

Tokyo, July 25-27, 2014

Sitemap

- LabPhon 14
- Satellite Events
- Links
- Contact
- About Laboratory Phonology

Back to home Back Top

Sitemap

10-2 Midori-cho, Tachikawa City, Tokyo, Japan 190-8561

LabPhon 14 - The 14th Conference on Laboratory Phonology

Tokyo, July 25-27, 2014

Satellite Events

- Pre-Conference Colloquium
- Satellite Workshop
- Call for Proposals

LabPhon 14 - The 14th Conference on Laboratory Phonology

Tokyo, July 25-27, 2014

Links

A Brief Sketch of Japanese Orthography and Phonology

- PDF file

NINJAL - National Institute for Japanese Language and Linguistics

- NINJAL Homepage

Phonology at NINJAL

- Phonological Characteristics of the Japanese Lexicon
- The Japanese Lexicon: A Rendaku Encyclopedia
- Conferences and Other Events

Project on Endangered Languages at NINJAL

- General Research for the Study and Conservation of Endangered Dialects in Japan (in Japanese)

Corpus Linguistics at NINJAL

- The KOTONOHA Project

Related Linguistic Societies

- The Acoustical Society of Japan
- The Linguistic Society of Japan
- The Phonetic Society of Japan
- The Phonological Society of Japan (in Japanese)

Association for Laboratory Phonology

- Link to the LabPhon website

- LabPhon 14
- Satellite Events
- Links
- Contact
- About Laboratory Phonology

Back to home | Back | Top

Sitemap

LabPhon 14 - The 14th Conference on Laboratory Phonology

Tokyo, July 25-27, 2014

Contact

Email Us

- **labphon14@ninjal.ac.jp**
- **If you have any questions, please let us know by email.**

10-2 Midori-cho, Tachikawa City, Tokyo, Japan 190-8561

LabPhon 14 - The 14th Conference on Laboratory Phonology

Tokyo, July 25-27, 2014

About Laboratory Phonology

Association for Laboratory Phonology

- Link to the LabPhon website

Previous LabPhon conferences

- LabPhon1 (1987), Columbus, OH, USA
- LabPhon2 (1989), Edinburgh, UK
- LabPhon3 (1991), Los Angeles, CA, USA
- LabPhon4 (1993), Oxford, UK
- LabPhon5 (1996), Evanston, IL, USA
- LabPhon6 (1998), York, UK
- LabPhon7 (2000), Nijmegen, The Netherlands
- LabPhon8 (2002), New Haven, CT, USA
- LabPhon9 (2004), Urbana, IL, USA
- LabPhon10 (2006), Paris, France
- LabPhon11 (2008), Wellington, New Zealand
- LabPhon12 (2010), Albuquerque, NM, USA
- LabPhon13 (2012), Stuttgart, Germany

LabPhon 14
Satellite Events
Links
Contact
About Laboratory Phonology

Back to home Back Top

Sitemap

10-2 Midori-cho, Tachikawa City, Tokyo, Japan 190-8561

LabPhon 14 - The 14th Conference on Laboratory Phonology

Tokyo, July 25-27, 2014

Call for Papers

- The Call for Papers has been closed.

Oral Presentations and Poster Presentations

- Abstracts are solicited for contributed papers for presentation as 20-minute oral contributions or as posters. Contributions relating to the conference themes are especially encouraged; there will also be sessions for non-thematic papers.

  - Abstract submission period: December 1, 2013 - January 31, 2014 (Honolulu time, UTC-10)
  - Notification of acceptance: March 15, 2014

Thematic and Non-thematic Sessions

- The overall theme for the conference is **"Laboratory Phonology beyond the Laboratory: Quantitative Analyses of Speech Produced outside the Phonetics Laboratory"**. Our goal is to bring together researchers from phonology, phonetics, and adjacent psycho- and neurosciences and to seek to advance these disciplines by encouraging the joint pursuit of interdisciplinary research questions. Specific topics that address this theme are the following:

  - *fieldwork-based studies, particularly on endangered languages/dialects*
  - *corpus-based approaches to spontaneous speech*
  - *acquisition of L1 phonology/prosody*

- Non-thematic sessions (both oral and poster) will include contributions to other topics of interest to the LabPhon community.

Financial Support

- We plan to provide financial support for at least 30 student presenters (either oral or poster presentations). If you are a student and the sole or first author of a presentation, you are eligible. The money available will cover lodging expenses for up to four nights in Tokyo. We will not be able to help with transportation expenses.

- Shortly after the abstract review process is completed, we will send more detailed information to the authors of all accepted abstracts.

Abstract Submission

- To submit an abstract, attach both the abstract itself (as both a Word and a PDF file) and a filled-out submission form (see below) to an e-mail message addressed to:

  labphon14@ninjal.ac.jp

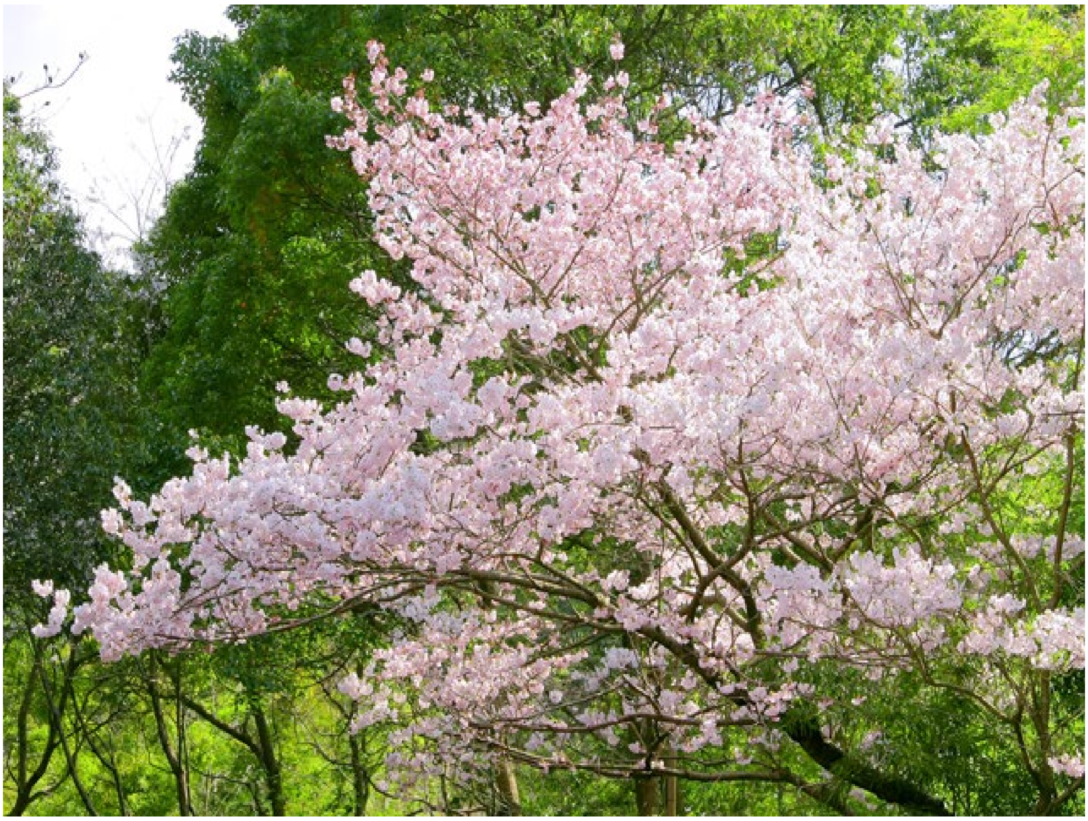
- The maximum length for the abstract is one page. This length restriction will be strictly enforced. A Word document template will be made available on this page soon. Please make sure that all special fonts and symbols are embedded in the PDF document and display properly.
- In addition, a submission form (see below) must accompany each abstract, attached it to the same e-mail message. The submission form can be either a Word file or a PDF file.
- One individual may be an author on no more than two submitted abstracts. All abstracts must be written in English. Presentations will be given in English. The abstracts will be evaluated anonymously by the scientific committee.
- The submission form asks you to identify which of the conference themes your paper relates to. If your paper does not relate to any of the themes, please select “non-thematic”. The submission form also asks you to specify whether you want to have your abstract considered for (1) either oral or poster presentation, (2) oral presentation only, or (3) poster presentation only. Please specify one of these three categories.

Abstract Template and Submission Form

- Abstract template
- Submission form

Publication of a Special Issue of Laboratory Phonology

- One issue of Laboratory Phonology is reserved for papers from LabPhon 14 (guest editors Haruo Kubozono, Kikuo Maekawa, and Timothy J. Vance). Owing to space limitations, the special issue will be restricted to papers that address the main theme of the conference "Laboratory Phonology beyond the Laboratory: Quantitative Analyses of Speech Produced outside the Phonetics Laboratory". Authors of papers in this category will be contacted personally. All authors whose papers are not considered for the special issue are encouraged to submit their manuscripts to the Laboratory Phonology journal as regular submissions.



LabPhon 14 - The 14th Conference on Laboratory Phonology

Tokyo, July 25-27, 2014

Updated Information

2013

LabPhon 14
Satellite Events
Links
Contact
About Laboratory Phonology

Back to home Back Top

Sitemap

10-2 Midori-cho, Tachikawa City, Tokyo, Japan 190-8561

Some cognitive factors behind vowel lengthening in spontaneous Japanese: A corpus-based study

Yasuharu Den
Faculty of Letters, Chiba University

Vowel length in continuous speech is determined by various factors. Previous studies in phonetic science and speech synthesis have identified the following factors, among others, that may affect the duration of a vowel: i) the inherent duration of the vowel, ii) the compensatory effect of the surrounding consonants, iii) the syntactic category of the word, iv) the position of the vowel in the word, the prosodic phrase, and the utterance, and v) the overall speech rate of the speaker. When we turn our attention to spontaneous discourse, however, we find other higher-level factors triggering the lengthening of vowels. Since in spontaneous discourse speakers need extra time to plan what to say next and to formulate it in an appropriate construction, they may suspend the ongoing utterance, stretching the speech segment being articulated, when they experience a severe cognitive load.

In this talk, I will investigate vowel lengthening in spontaneous Japanese, based on a quantitative analysis of a large-scale corpus. In particular, I focus on three locations at which vowel lengthening frequently occurs: the final vowel i) of an utterance preface (filler or discourse marker) such as *e* and *de*, ii) at the end of a preceding utterance, and iii) of a topic phrase marked by the particle *wa*. The following example illustrates these three sites:

... suru-n-desu-keredo-mo:	(1.1)	e:	(0.3)	saiaku-na-no-wa:	zieetai-ni ...
do-N-POL-yet		um		worse-COP-N-TOP	SDF-DAT
Preceding utterance		Preface		Topic	

A series of our studies have shown that the durations of vowels at these locations are positively correlated with the length of the following utterance (*zieetai-ni ...*) and that they are also affected by the strength of connectivity between the preceding and the current utterances. By using novel statistical modeling and incorporating the above mentioned lower-level factors into the models, I will demonstrate how vowel lengthening in spontaneous Japanese is governed by cognitive factors, more precisely, by planning of utterance and discourse.

Some cognitive factors behind vowel lengthening in spontaneous Japanese: A corpus-based study

Yasuharu Den

Faculty of Letters
Chiba University

LabPhon14@NINJAL

Spontaneous speech

- Ano: sono-ko-wa nihon-ni ki-te: (0.9) ma: (0.3) ano uti: watasi-no it-teru tokoro-wa nihon-go: (0.5) -ga^ (0.3) n nihon-go-gaku-tte-yu-no-o yat-teru tokoro-na-node (0.3) ma: ano: nihon-ni kuru mae-mo moo nihon-go perapera-da-si: (CSJ:D01F0023)

Spontaneous speech

- Ano: sono-ko-wa nihon-ni ki-te: (0.9) ma: (0.3) ano uti: watasi-no it-teru tokoro-wa nihon-go: (0.5) -ga^ (0.3) n nihon-go-gaku-tte-yu-no-o yat-teru tokoro-na-node (0.3) ma: ano: nihon-ni kuru mae-mo moo nihon-go perapera-da-si: (CSJ:D01F0023)
- Various sorts of disfluencies:
  - Fillers, suspensions, repairs, repetitions, & segment lengthening

Today's topic

- Vowel lengthening (aka prolongation)
  - Non-lexical stretching of speech segments
  - Occurring everywhere in an utterance
- Examples from the previous excerpt
  - End of fillers: *ano:*, *ma:*
  - End of phrases: *uti:*, *nihon-go:*
  - End of clauses: *ki-te:*, *perapera-da-si:*
  - Over 90% of lengthening in Japanese occurs at the end of words (Den, 2003).
- Q: What factors behind vowel lengthening?

Outline

1. Background
2. Data, annotation, and statistical methods
3. Analysis 1: Lengthening at the beginning of utterances: Fillers and conjunctions
4. Analysis 2: Lengthening at the beginning of utterances: *Wa*-marked phrases
5. Analysis 3: Lengthening at the end of utterances
6. Discussion

BACKGROUND

Vowel lengthening

- Vowel lengthening has been studied in phonology, phonetics, and speech synthesis research.
- Various factors
  - Final lengthening at various levels (Klatt, 1975)
    - Word < Accentual phrase < Intonation phrase
  - Pre-pausal lengthening
  - Compensatory effect of the surrounding phonemes, e.g. mora-timed rhythm in Japanese
  - Simultaneous lengthening of successive phonemes within a syllable (Campbell & Isard, 1991)

Factors used in speech synthesis

- These factors, among others, have been applied to speech synthesis (Kaiki et al., 1990):
  - compensatory effect of the surrounding consonants
  - position of the vowel in the word, the prosodic phrase, and the utterance
  - presence of the following pause
  - syntactic category of the word
  - inherent duration of the vowel
  - overall speech rate of the speaker

Prolongation in spontaneous speech

- Swedish (Eklund, 2001)
  - focused on phonological & morpho-syntactic factors such as phone type, position in the word, lexical factors, and word class
- Japanese (Den, 2003)
  - found some strategies in prolonging speech segment used by Japanese speakers
- Mandarin (Lee et al., 2004)
  - took functional difference into account such as hesitation, emphasis, and feedback

Possible other factors

- Only linguistic factors so far
- In spontaneous speech, other factors may affect vowel lengthening.
- Planning load = Cognitive factor
  - In spontaneous discourse, speech planning is done on the fly.
  - Speakers may take extra time in planning complex utterances.
  - On these occasions, they signal their problems in the form of disfluencies (Clark, 2002):
    - Fillers, repetitions, repairs, and segment lengthening

Previous findings in my studies

- Utterance initial fillers and conjunctions
  - Filler *e*: positively correlated with the duration of the utterance (under some conditions) (Watanabe & Den, 2010)
  - Conjunction *de*: no such correlation (Den, 2009; Watanabe & Den, 2010).
- Utterance initial *wa*-marked topic phrases
  - *Wa*: positively correlated with the duration of the rest of the utterance (Watanabe & Den, 2010; Den & Nakagawa 2013).
- End of clauses
  - Final mora: positively correlated with the duration of the next clause (within an utterance) (Koiso & Den, 2013).

Problems of the previous studies

- Phonological and morpho-syntactic factors were not fully considered (nor controlled).
- The cause-effect relationship, i.e. which is dep. variable and which is indep. variable, was not consistent across studies.
- The relationship among lengthening at difference places was not investigated.

- In this talk, I solves some of these problems.

DATA, ANNOTATION, AND STATISTICAL METHODS

Agenda for methodology

- The study of spontaneous speech
  - Difficult to apply experimental methods
  - Important to investigate natural speech data
- But, natural speech data is messy.
- Requirements for dealing with natural speech data
  - Big amount of data

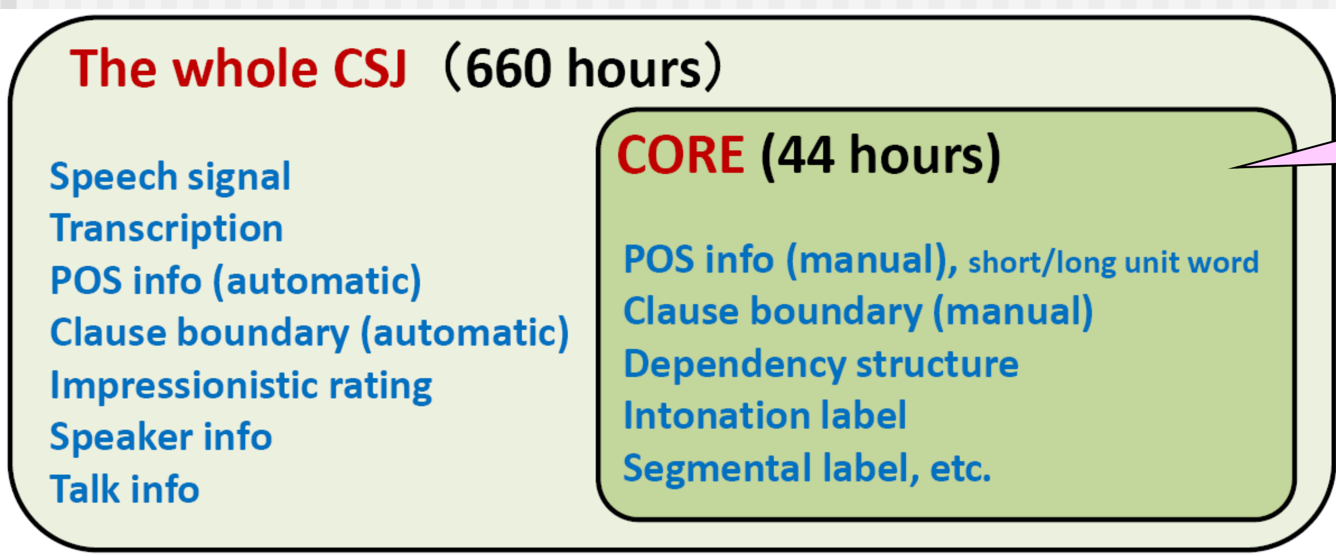
    -> *Corpus of Spontaneous Japanese*
  - Control of confounding variables

    -> Data selection & covariates
  - Adequate statistical method

    -> Mixed-effects model

Data

- *Corpus of Spontaneous Japanese* (CSJ)
  - Large-scale corpus of spontaneous Japanese, developed by NINJAL
  - Mostly monologs (625 hours)
    - Academic presentations and speech on everyday topics



Data analyzed
107 talks
20 hours
9.8K clause units
230K words

Annotation

- Linguistic annotations at various levels
  - Phonemes
    - Starting and ending times, their uncertainty, devoicing, etc.
  - (Long- & short-unit) words
    - Part-of-speech, conjugation form, dictionary form, etc.
  - Accentual phrases
    - Break indices and boundary tones (based on X-JToBI)
  - Bunsetsu phrases
    - Dependency structures
  - Clause units (regarded as utterances)
    - Clause boundary (CB) types
- Compiled in a relational database (Koiso et al., 2014)

CB types

- Absolute (AB)
  - sentence ending in usual sense

  E.g. Tokyo-ni iki-masu
  Tokyo-DAT go-POL
  I will go to Tokyo.

- Strong (SB)
  - clause boundary with coordinate particle

  E.g. Tokyo-ni iki-masu-ga
  Tokyo-DAT go-POL-but
  I will go to Tokyo, but …

- It is sometimes better to consider other types of clauses and phrases as independent utterances.
  - Weak (WB)

  E.g. Tokyo-ni iku-node
  Tokyo-DAT go-because
  Because I go to Tokyo.

  - Non-Clausal (NCB)

  E.g. Tokyo-ni
  Tokyo-DAT
  To Tokyo

AB > SB > WB > NCB

Variables

- Schematic representation of the utterance

… suru-n-desu-keredo-mo: (1.1) e: (0.3) saiaku-na-no-wa: (1.0) zieetai-ni …
… do-N-POL-yet um horrible-COP-N-TOP SDF-DAT …

3. Preceding utterance — 1. Preface — 2. Topic — Body

  - Preface and topic may be missing.
- Dependent variable
  - Duration of the final vowel
- Independent variables
  - CB type of the preceding utterance
  - Duration of the body

Degree of disjuncture between utterances

Complexity of the utterance

Data selection

- Use only reliable data
  - E.g. exclude cases where the phoneme boundaries are uncertain
- Use only major categories
  - E.g. focus on frequent preface items such as filler *e* and conjunction *de*
- Use only simple cases
  - E.g. focus on topics with the simplest structure, i.e. Noun/Pronoun-*wa*

Statistical analysis

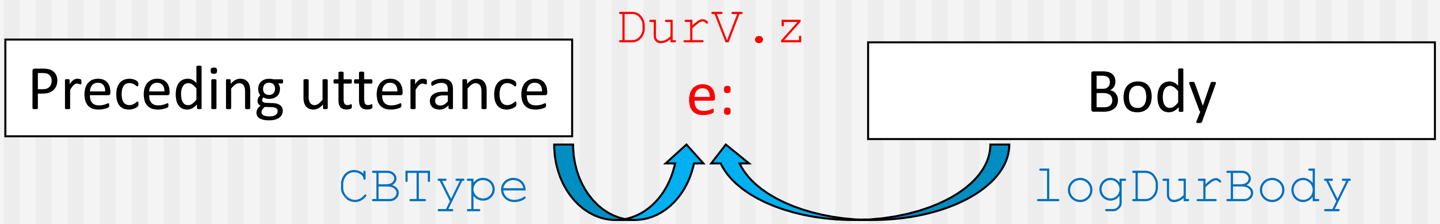
- To consider the inherent duration of the vowel and the overall speech rate of the speaker
  - z-score transformation (cf. Campbell & Isard, 1991)

$$Dur_z = \frac{\log(Dur_{raw}) - mean}{sd}$$

    - mean/sd: calculated for each phoneme and each speaker
- To consider the hierarchical structure of the corpus data, i.e. clusters according to speakers
  - mixed-effects models (Baayen, 2008)
  - random intercept for speakers and word forms

ANALYSIS 1:
FILLERS AND CONJUNCTIONS

Method



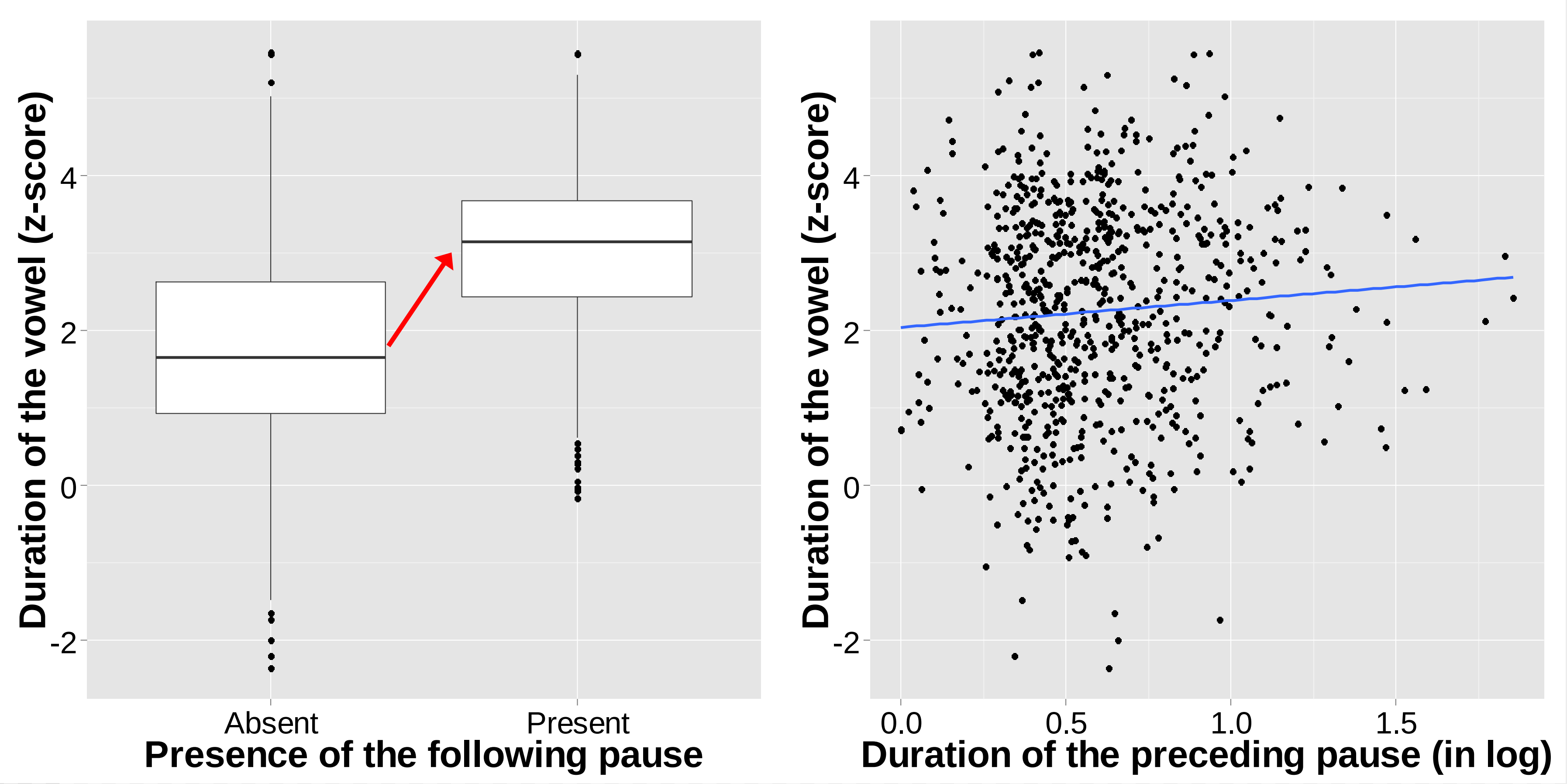
- **Data selection**
  1. Limited to four major categories (> 66%)
     - Filler *e*, Filler *ma*, Filler *ano*, & Conjunction *de*
  2. Excluding uncertain phoneme boundaries and non-canonical pronunciation
  3. Limited to simple patterns, i.e. utterance initial fillers/conjunctions followed by no other preface items
- **Data analyzed**

Filler *e*	Filler *ma*	Filler *ano*	Conj *de*
761	615	353	839

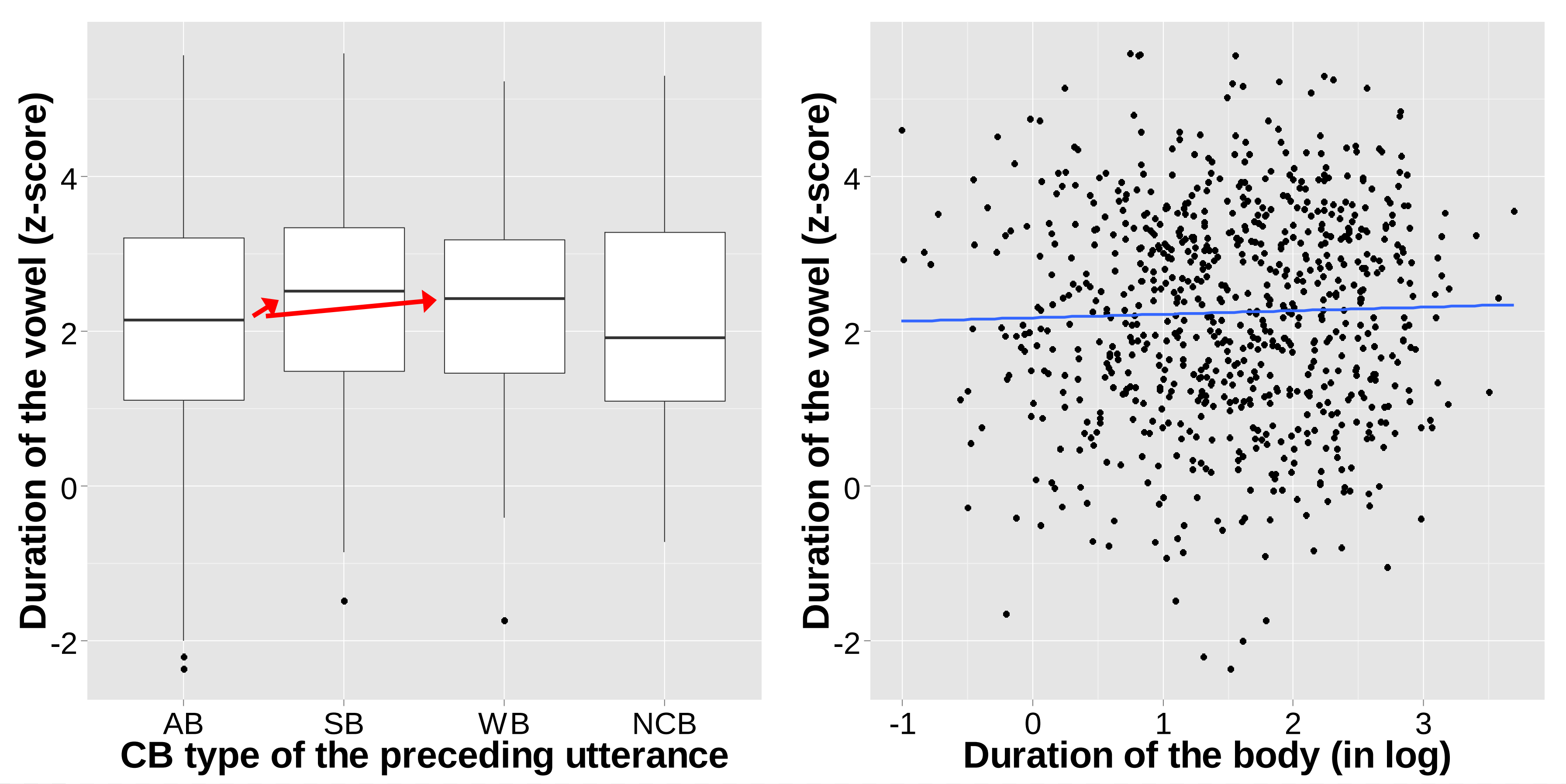
Method (cnt'd)

- Variables
  - Independent variables
    - CB type of the preceding clause (`CBType`)
    - Duration of the body (in log) (`logDurBody`)
    - Their interaction (not significant, removed)
  - Covariates
    - Duration of the preceding consonant (`DurC.z`)
    - Presence of the following pause (`ifFolPause`)
    - Duration of the preceding pause (in log) (`logDurPrePause`)
    - Presence of the topic (`ifTopic`)
  - Random effects
    - Intercept for speakers
- Parameter estimation
  - Maximal Likelihood Estimation by `lmer` of R language
  - *p*-values obtained by likelihood ratio tests

Results: *e* vs covariates



Results: *e* vs indep. variables



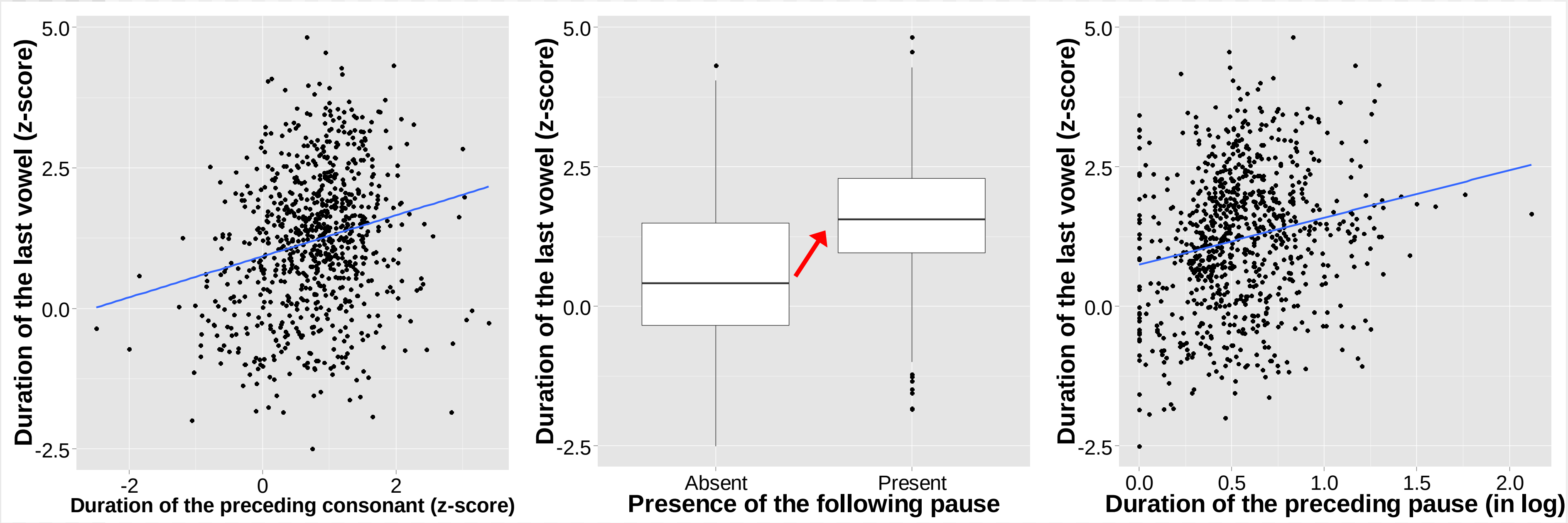
Results: Mixed-effects model

- All covariates had a significant effect.
- The CB type had a significant effect (AB < SB, WB), but the body duration did not.

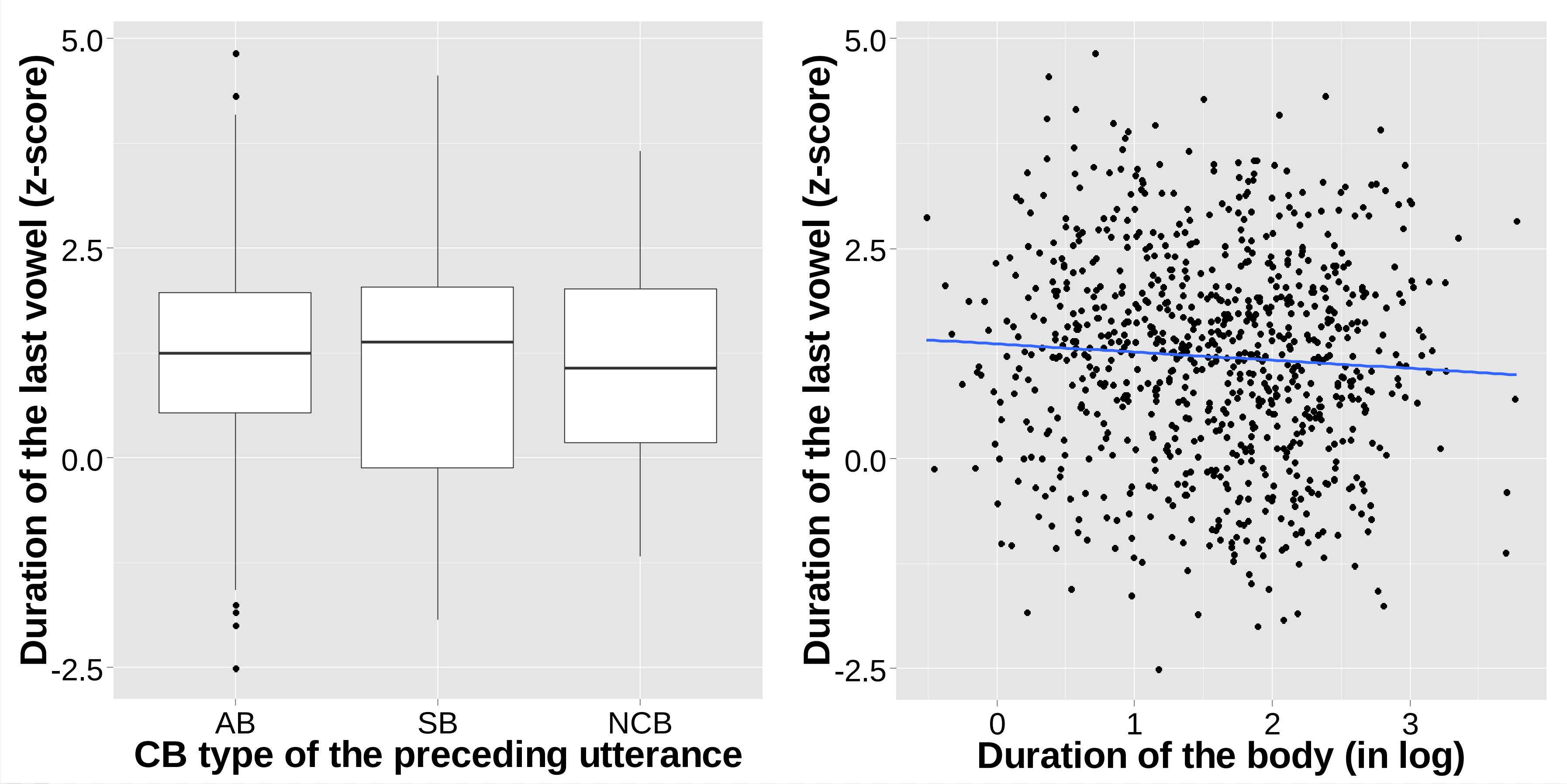
	Coef.	SE	*t* value	*p* value
CBType=SB	.356	.100	3.55	.002
CBType=WB	.332	.137	2.43	
CBType=NCB	.053	.174	.31	
logDurBody	.066	.049	1.35	.177
ifFolPause	1.214	.087	13.88	.000
logDurPrecPause	.352	.174	2.03	.043
ifTopic	-.150	.106	-1.42	.157

$\sigma = 1.07$, $\sigma_{Speaker} = .65$

Results: *de* vs covariates



Results: *de* vs indep. variables



Results: Mixed-effects model

- All covariates had significant effects.
- No significant effects of the CB type or the body duration

	Coef.	SE	*t* value	*p* value
CBType=SB	-.055	.088	-.62	.804
CBType=NCB	.028	.209	.13	
logDurBody	-.065	.046	-1.43	.154
DurC.z	.345	.051	6.72	.000
ifFolPause	.950	.076	12.47	.000
logDurPrecPause	.753	.142	5.30	.000
ifTopic	-.173	.093	-1.88	.062

$\sigma = .98$, $\sigma_{Speaker} = .47$

Summary of the results

- All phonological covariates had significant effects.
- The effect of the CB type was significant in fillers *e* and *ma*, but the effect of the body duration was significant only in filler *ma*.

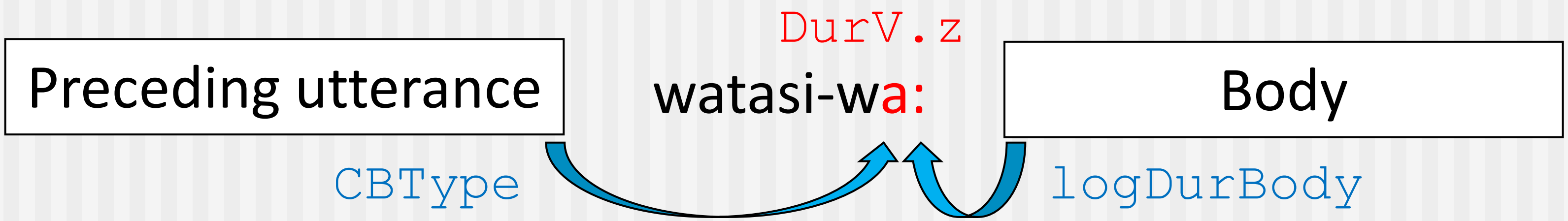
	Filler *e*	Filler *ma*	Filler *ano*	Conj *de*
CBType	AB < SB, WB	AB < SB, WB	ns	ns
logDurBody	ns	+	ns	ns
DurC.z		+	+	+
ifFolPause	+	+	+	+
logDurPrecPause	+	+	ns	+
ifTopic	ns	ns	ns	ns

Summary of Analysis 1

- Lengthening of the last vowel in utterance initial fillers and conjunctions is
  - consistently affected by phonological factors:
    - the duration of the preceding consonant
    - the presence of the following pause
  - but not always affected by cognitive factors
- Among fillers, *ma* is most affected by cognitive factors, *e* next, and *ano* least.
- Conjunction *de* seems independent of cognitive factors.

ANALYSIS 2:
WA-MARKED PHRASES

Method



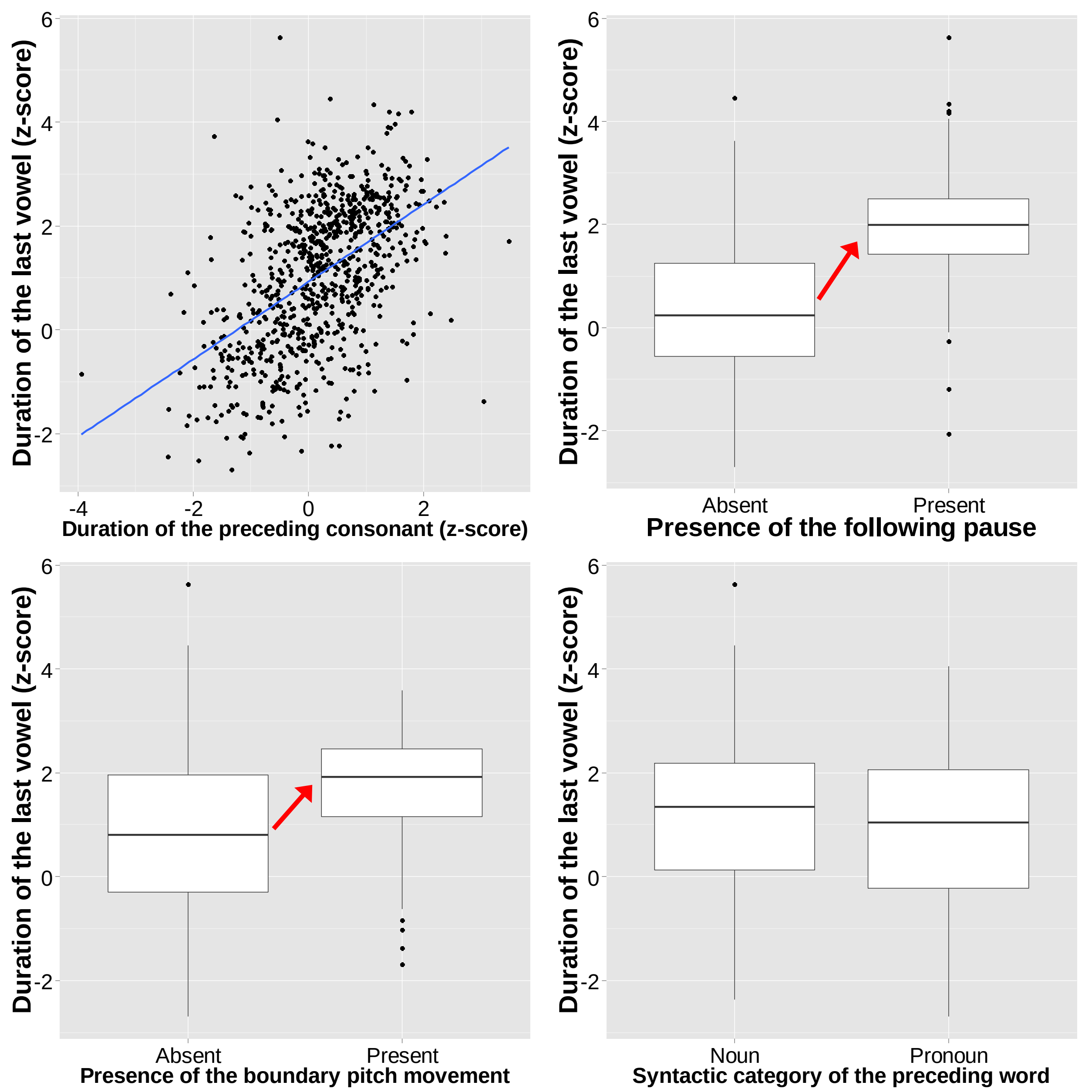
- Data selection
  1. Limited to those starting with noun or pronoun (> 97%)
  2. Excluding uncertain phoneme boundaries and non-canonical pronunciation
  3. Limited to simple phrases, i.e. Noun/Pronoun-*wa* (including Noun/Pronoun-*toiu-no-wa* and the like)
- Data analyzed

Noun-*wa*	Pronoun-*wa*
464	337

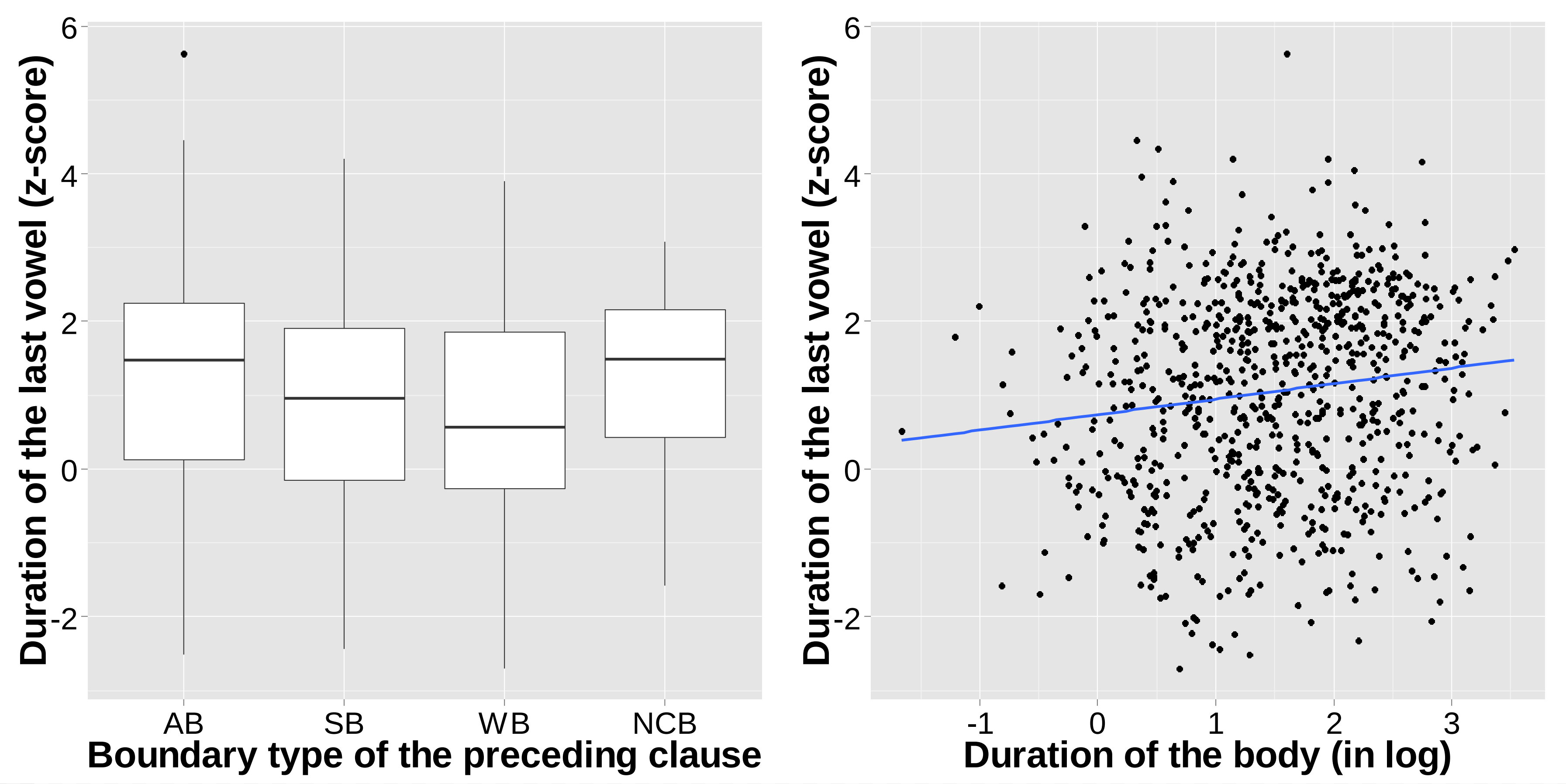
Method (cnt'd)

- Variables
  - Independent variables
    - CB type of the preceding clause (`CBType`)
    - Duration of the body (in log) (`logDurBody`)
    - Their interaction (not significant, removed)
  - Covariates
    - Duration of the preceding consonant (`DurC.z`)
    - Presence of the following pause (`ifFolPause`)
    - Presence of the boundary pitch movement (`ifBPM`)
    - Syntactic category of the preceding word (`ifPrePronoun`)
    - Presence of the preface (`ifPreface`)
  - Random effects
    - Intercept for speakers
    - Intercept for word forms (not significant, removed)

Results:
Wa vs covariates



Results: *Wa* vs indep. variables



Results: Mixed-effects model

- All phonological covariates had significant effects.
- The effect of the body duration was also significant.

	Coef.	SE	*t* value	*p* value
CBType=SB	-.172	.083	-2.08	
CBType=WB	-.270	.130	-2.08	.079
CBType=NCB	-.028	.148	-.19	
logDurBody	.092	.041	2.23	.026
DurC.z	.472	.042	11.34	.000
ifFolPause	1.148	.079	14.54	.000
ifBPM	.346	.091	3.80	.000
ifPrePronoun	-.065	.075	-.88	.381
ifPreface	.037	.077	.48	.629

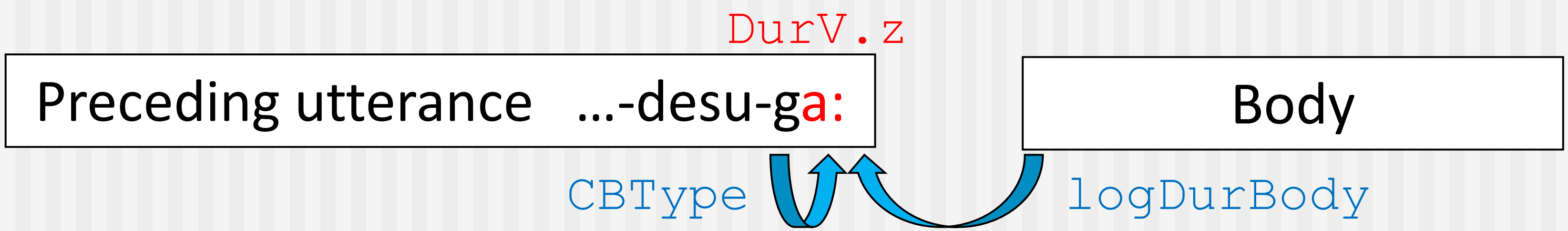
$\sigma = .97,\ \sigma_{\text{Speaker}} = .31$

Summary of Analysis 2

- Lengthening of the vowel *a* of *wa* in utterance initial topic phrases is
  - affected by phonological factors:
    - the duration of the preceding consonant
    - the presence of the following pause
    - the presence of the boundary pitch movement
  - and also affected by some cognitive factor, i.e. the complexity of (the rest of) the utterance

ANALYSIS 3:
END OF UTTERANCE

Method



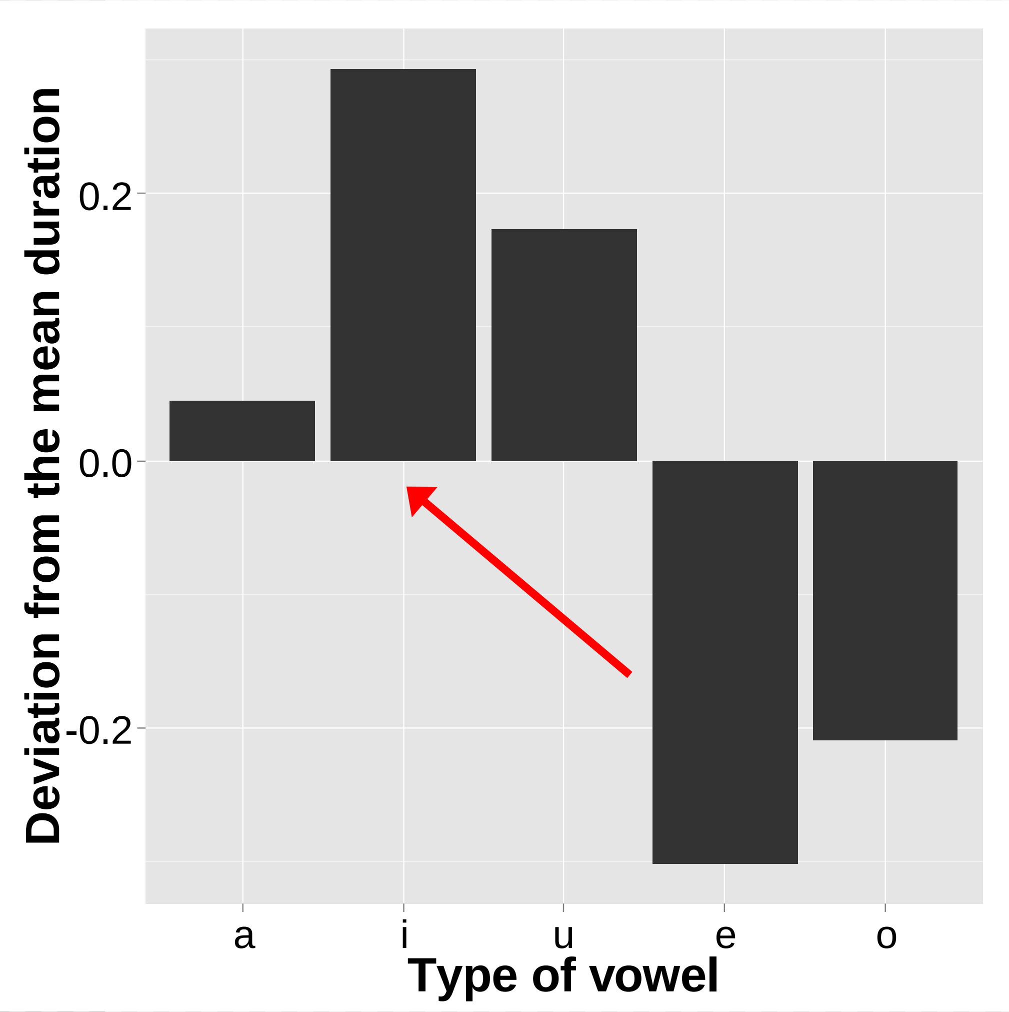
- Data selection
  1. Limited to those ending with particle or auxiliary verb (> 94%)
  2. Excluding uncertain phoneme boundaries, non-canonical pronunciation, and devoiced vowels
  3. Limited to those coincident with AP boundary
- Data analyzed

AB	SB	WB	NCB
2005	2940	738	253

Method (cnt'd)

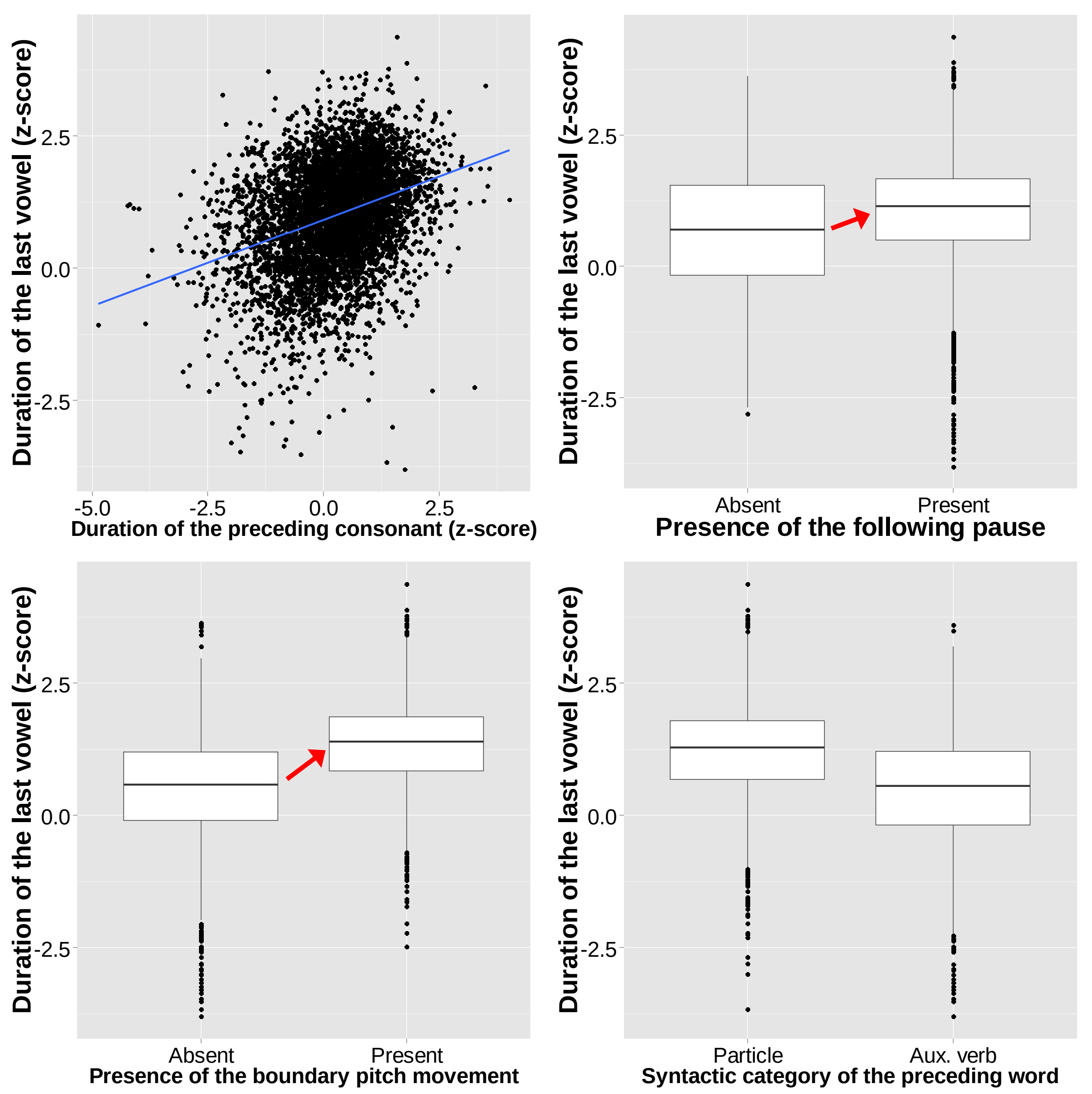
- Variables
  - Independent variables
    - CB type of the utterance (`CBType`)
    - Duration of the body (in log) (`logDurBody`)
    - Their interaction
  - Covariates
    - Type of the vowel (`VEntity`)
    - Duration of the preceding consonant (`DurC.z`)
    - Presence of the following pause (`ifFolPause`)
    - Presence of the boundary pitch movement (`ifBPM`)
    - Syntactic category of the word (`ifAuxV`)
    - Presence of the preface (`ifPreface`)
    - Presence of the topic (`ifTopic`)
  - Random effects
    - Intercept for speakers
    - Intercept for word form

Results: End of utt. vs vowel type

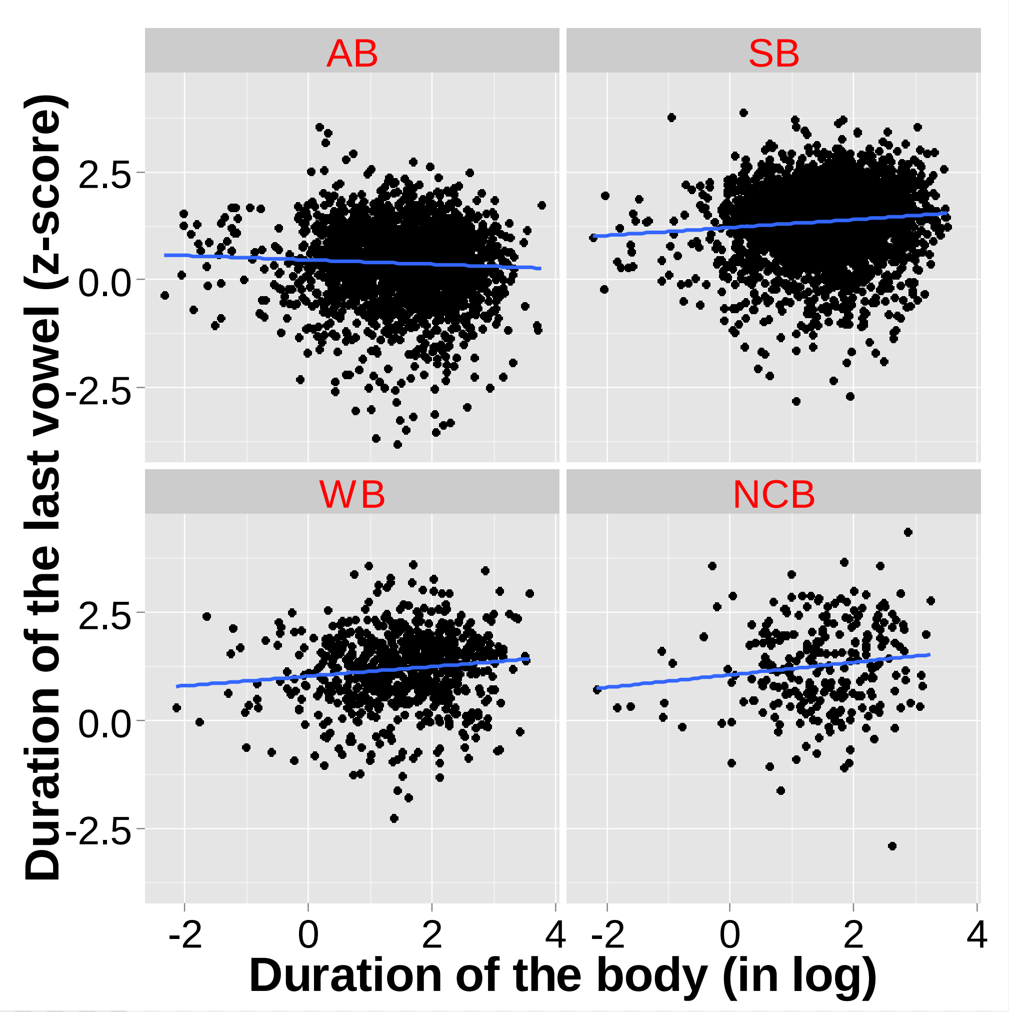


- *a, i, & u* were longer than *e* & *o* at the end of utterance (on z-score scale).

Results: End of utt. vs covariates



Results: Mixed-effects model

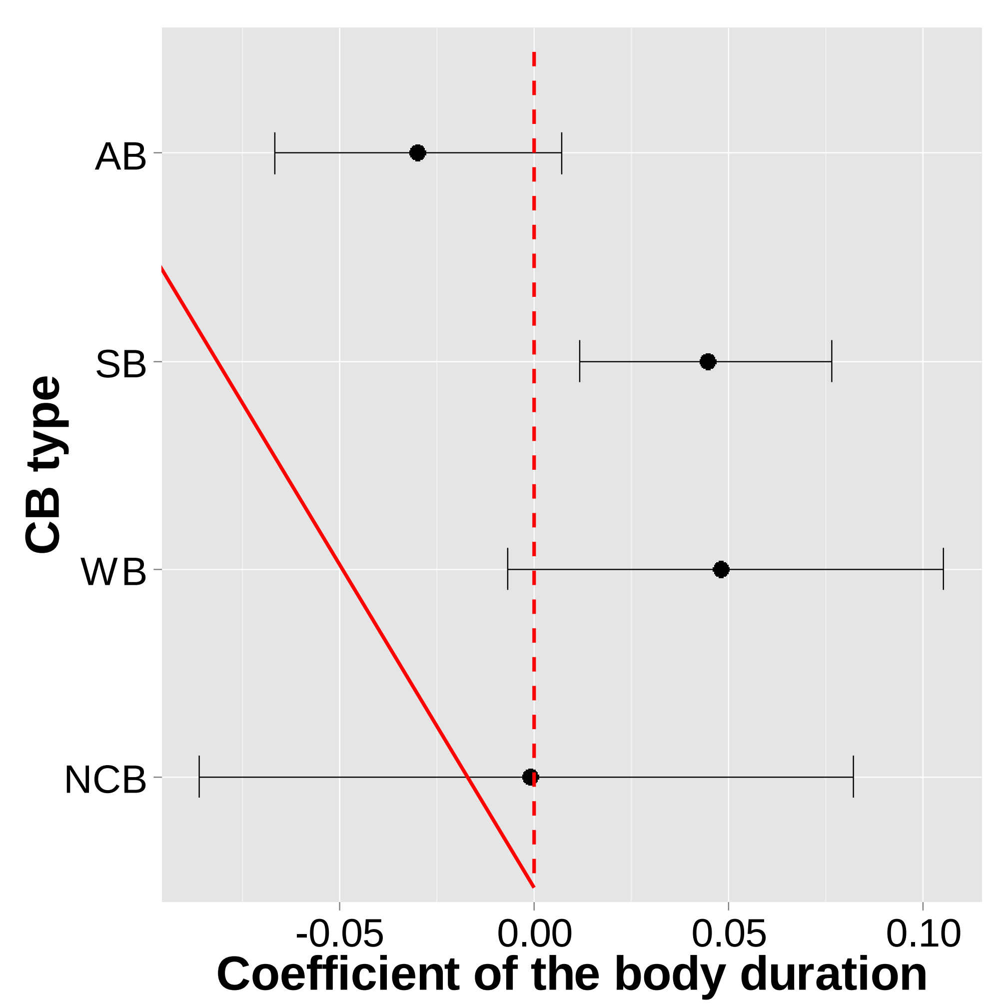


- All phonological covariates had significant effects.
- Significant interaction between the CB type and the body duration (p = .002 by LRT)

	Coef.	SE	*t* value	*p* value
`VEntity`				.016
`DurC.z`	.112	.012	9.60	.000
`ifFolPause`	.460	.036	12.95	.000
`ifBPM`	.553	.026	21.44	.000
`ifAuxV`	-.082	.084	-.98	.335
`ifPreface`	.019	.021	.92	.360
`ifTopic`	-.018	.026	-.70	.484

$\sigma = .72$, $\sigma_{Speaker} = .23$, $\sigma_{Orth} = .23$

Results: Coefs. of body duration



- To obtain precise estimates of the coefficients of the body duration for each CB type, we applied MCMC technique using JAGS language (Kruschke, 2011)
- The body duration had a significant positive coefficient only when the CB type was the strong boundary.

Summary of Analysis 3

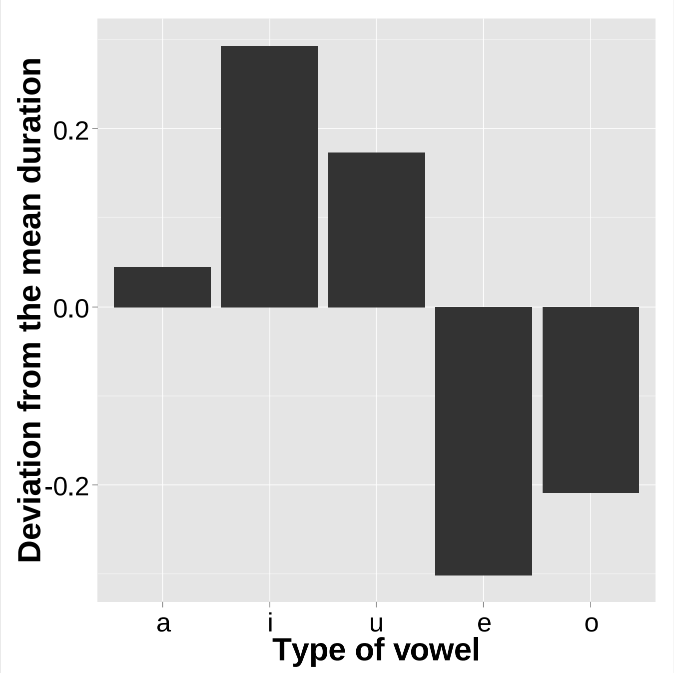
- Lengthening of the last vowels of utterances is
  - affected by phonological factors:
    - the duration of the preceding consonant
    - the presence of the following pause
    - the presence of the boundary pitch movement
    - the type of the vowel
  - and also affected by the complexity of the following utterance under some conditions, i.e. when ending with a strong boundary

DISCUSSION

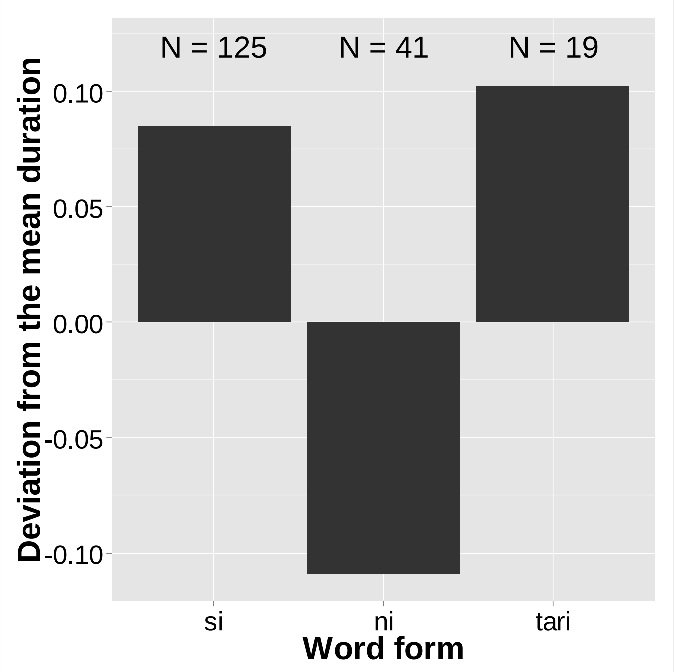
Phonological factors

- At all places, our phonological factors have reliable effects, i.e. vowel lengthening is enhanced when
  - the duration of the preceding consonant is longer;
  - the vowel is followed by a pause; and
  - the vowel bears boundary pitch movement
- The effect of the preceding consonant is supplementary rather than compensatory, suggesting that the entire mora, not just the vowel, is lengthened.

Effect of vowel type



- The degree of lengthening is affected by the vowel type.
  - *i* has the strongest effect, although its inherent duration is short.



- This effect is attributed mainly to a particular lexical item.
  - coordinate particle *si*, which appears at strong boundaries

Complexity effect

- The duration of the utterance body, i.e. the complexity, sometimes affects vowel lengthening.

	Filler *ma*	Topic marker *wa*	Last mora at SB
Coef. of `logDurBody`	.157	.092	.048

- The significant effect found at the last mora of the preceding utterance with a strong boundary may suggest that Japanese speakers use early signal for upcoming troubles.
- But, these coefficients are rather small compared with those of the covariates, e.g. 1.053 for the following pause and .342 for BPM in the case of *wa*.

Disjuncture effect

- The CB type, i.e. the degree of disjuncture between utterances, is sometimes relevant.
  - Filler *e* is longer at SB & WB than at AB.
  - The complexity effect on the last vowel of the preceding utterance is significant only when the preceding utterance ends with SB.
- Two possible explanation for weaker effect at AB
  1. The data for AB is distorted.
     - In analysis 3, the data selection step removed 50% of the data for AB (due to devoicing in *desu* & *masu*) but only 15-25% for SB & WB.
  2. Some different cognitive process is involved at AB.
     - E.g. discourse-level planning

Summary

- Vowel lengthening in spontaneous Japanese
  - Phonological factors
  - Cognitive factors
- Further Q: Relationship among lengthening at different places?
  - Complementary, supplementary, or independent?
  - Related to different functions?
- Ready to go out of laboratories!
  - Adequate corpora and analytic methods

Thank you for your kind attention
and
Let’s enjoy fireworks!

Special thanks to:
Michiko Watanabe
Natsuko Nakagawa
and Hanae Koiso

On establishing the existence of word stress

Carlos Gussenhoven

Radboud University Nijmegen

A language has word stress if a syllable-based culminative and obligatory prominence feature is part of the phonology of words (Hyman 2006). This definition excludes languages with a mora-based culminative and obligatory tone, like Kinga, languages with an obligatory phrase-based syllabic pitch accent, like French, and languages with non-obligatory syllable-based prominence, like Japanese.

A criterion not listed by Hyman is phonetic prominence. I will discuss data from a number of languages in which the relation between phonetic salience and stress is unexpected, confirming that phonetic salience measures do not define word stress.

In Ambonese Malay, a language without vowel quantity, minimal pairs like [barat] 'West' – [băraat] 'heavy' suggest that the language has word stress, but it is hard to make a case for its existence. The salient peaks would appear to be best analyzed as due to phrase-boundary melodies that remain floating. This position will be argued for on the basis of a peak alignment study. Second, while the status of word stress in varieties of Tamazight is ambiguous at best, in the Zuara variety penultimate stress is a regular feature of words, even those that have a voiceless obstruent in the rime of the penultimate syllable, like [a.ˈsq.qad] 'flail'. This position will be defended on the basis of acoustic measurements in questions and statements. Third, Standard Nigerian English has tonal structures which reflect the position of the word stress in British English. However, while other new varieties of English with tonal substrates, like Cantonese English, apparently lack word stress, Standard Nigerian English distinguishes words with initial and peninitial stress by means of duration as well as pitch. Its word prosodic structure will be argued for on the basis of acceptability judgements of sentential stimuli in which f0 has been manipulated.

An operational definition of word stress may be provided by the 'stress deafness' paradigm of Peperkamp & Dupoux and colleagues: if listeners perform poorly on reproducing the presentation order of series of stimuli that minimally differ in the position of phonetic prominence ([minú] – [mínu], …), the language doesn't have word stress. We ran a 'stress deafness' experiment to see whether Persian, which has generally been described as having contrastive word stress or accent, passes the 'stress' criterion, using languages that uncontroversially have lexical stress or accent and languages that don't as upper and lower baselines. The results suggest that the 'stress deafness' test discriminates between lexical and postlexical stress or accent, and that the reason that Persian listeners are 'stress deaf' is that their accent distinctions arise postlexically.

On establishing the existence of word stress

Carlos Gussenhoven

RADBOUD UNIVERSITEIT NIJMEGEN

14th Conference on Laboratory Phonology

NINJAL, Tachikawa

July 25-27, 2014

Structure of talk

- Stress definitions
- Nigerian English has tone and stress (Acceptability experiment, production experiment)
- Ambonese Malay has no stress (Production experiment).
- Zwara Tamazight has stress (Preliminary production data).
- Persian has no lexically marked word prosody.

Definitions of stress

- Obligatory and culminative syllabic prominence feature of words (Hyman 2006, 2012)
- Word based prominence realized by phonetic variables other than just f0 (‘stress accent’, Beckman 1986)
- Feature of a language whose speakers are not ‘stress-deaf’ (Peperkamp & Dupoux 2002)

Frequent additional properties: Predictability and privilege

- Predictable location (there's always a default location).
- Edge-orientation.
- Phonologized segmental correlates:
  - Weight to Stress Principle (Dutch)
  - Stress to Weight Principle (Dutch)
  - Restricted unstressed vowel set (Russian, English).
- Phonologized prosodic correlates:
  - Location of tone contrasts (Swedish)
  - Location of intonational pitch accents (Swedish, English).

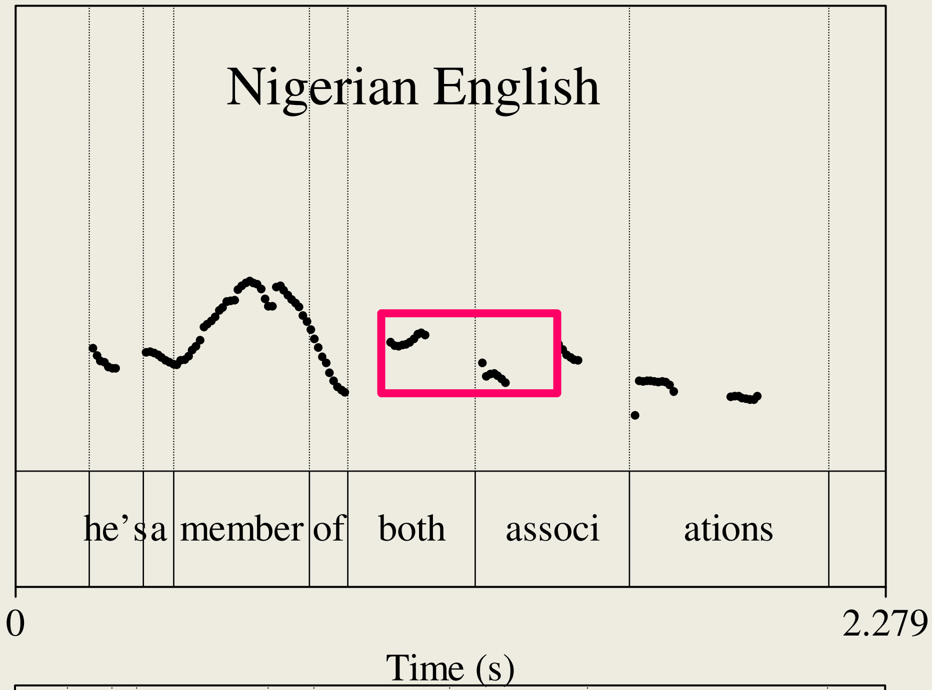
'Stress deafness'

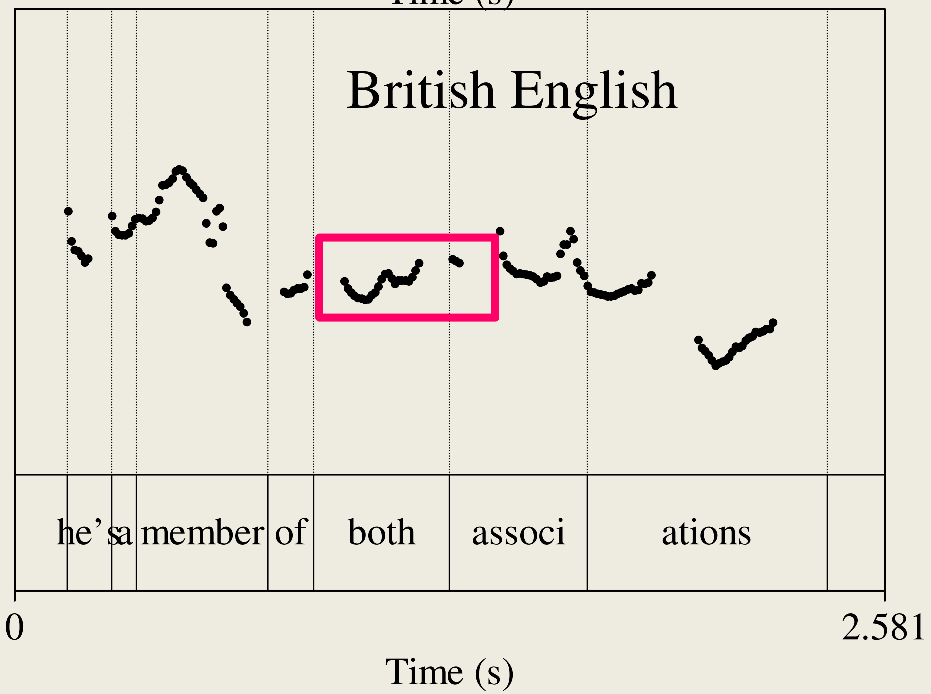
- Auditory stimuli: series of alternating stress patterns (númi – numí) (3, 4 or 5), followed by a distractor stimulus.
- Task: give the order of the stress patterns in the series.
- Baseline: phoneme differences (múni – múki)

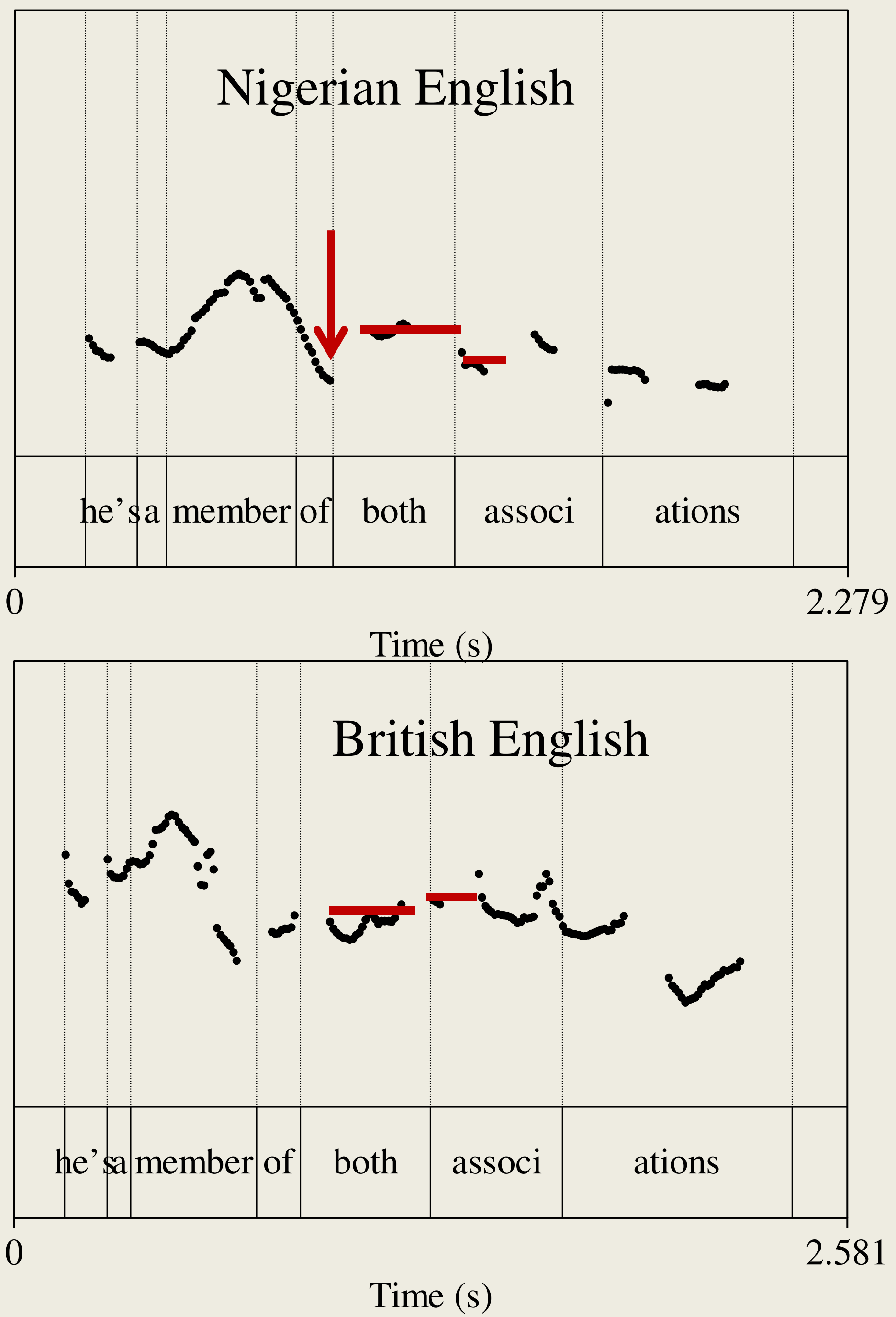
(Peperkamp & Dupoux [2002] and following work)

Nigerian English

- Nigerian English has many tonal substrate languages, and has become standardized (*Standard Nigerian English*, Eka 1985).


Nigerian English
he's a member of both associ ations
0
2.279
Time (s)


British English
he's a member of both associ ations
0
2.581
Time (s)



Nigerian English
he’s
a
member
of
both
associ
ations
0
2.279
Time (s)
British English
he’s
a
member
of
both
associ
ations
0
2.581
Time (s)

2010 acceptability experiment

- Is pitch drop at word boundary obligatory?
- Are function words L?

- Twelve words with peninitial stress.
- Final position in sentence.
- Function word in IP-medial position.
- 20 BrE and 20 Nigerian listeners.

Gussenhoven & Udofot (2010)

Tasks

- SNE listeners: These are educated speakers of Nigerian English (broadcasting, university teaching). Rate acceptability.

- BrE listeners: These are Nigerian students who took a course in British English intonation. Rate acceptability.

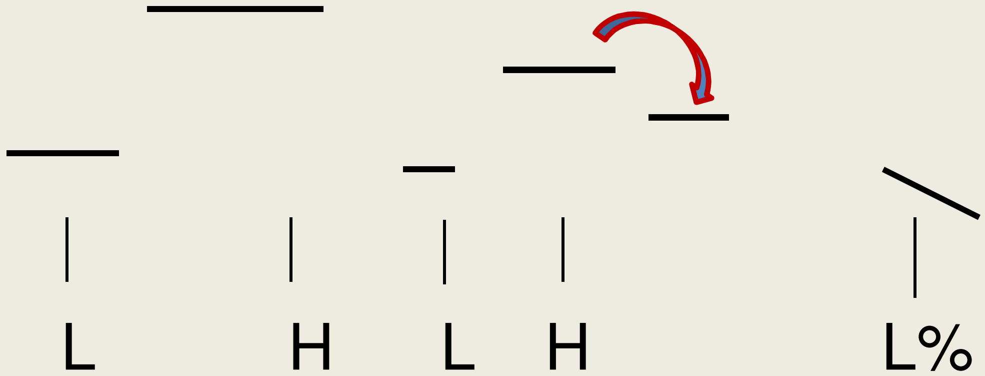
Results

- Yes, in NigE pitch is lowered at the word boundary, and
- yes, function words are L.
- Neither of which are BrE.

Gussenhoven & Udofot (2010)

What is that first syllable: !H or L?

They grow the best bananas



Standard Nigerian English

- Lexical words: H
- Function words: L
- Declarative: L%
- Interrogative: H%
- Downstep (IP): H → [!H] / H ... ___

(Lexical words and Function words: Udofot 2007)

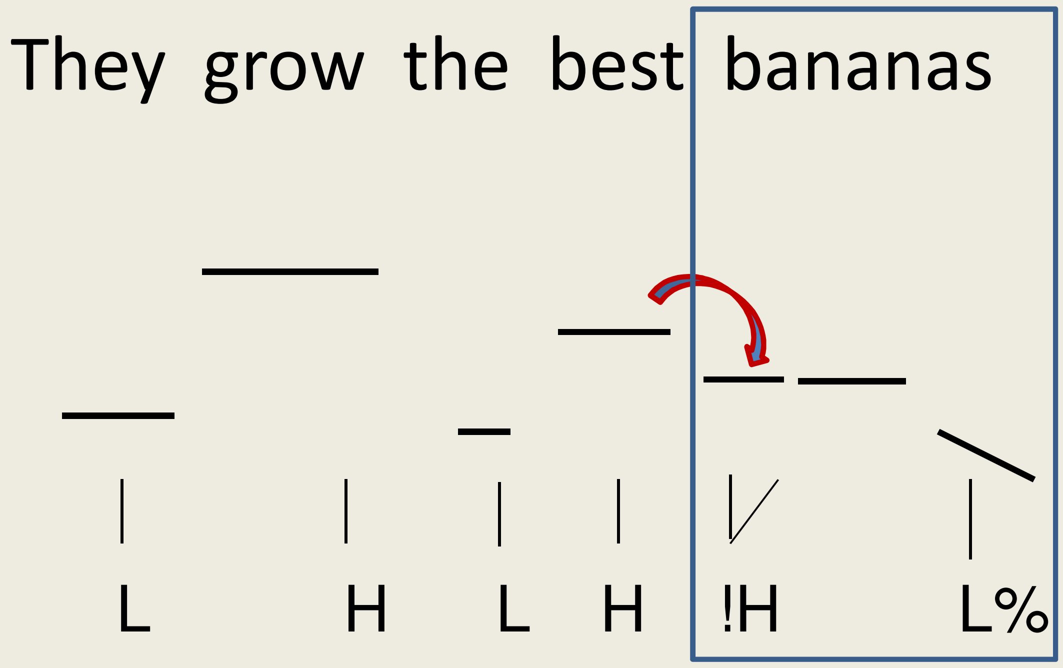
Standard Nigerian English

- Lexical stress: (L)H
- Function words: L

  } Gut 2005
- Declarative: L%
- Interrogative: H%
- Downstep (IP): H → [!H] / H ... ___

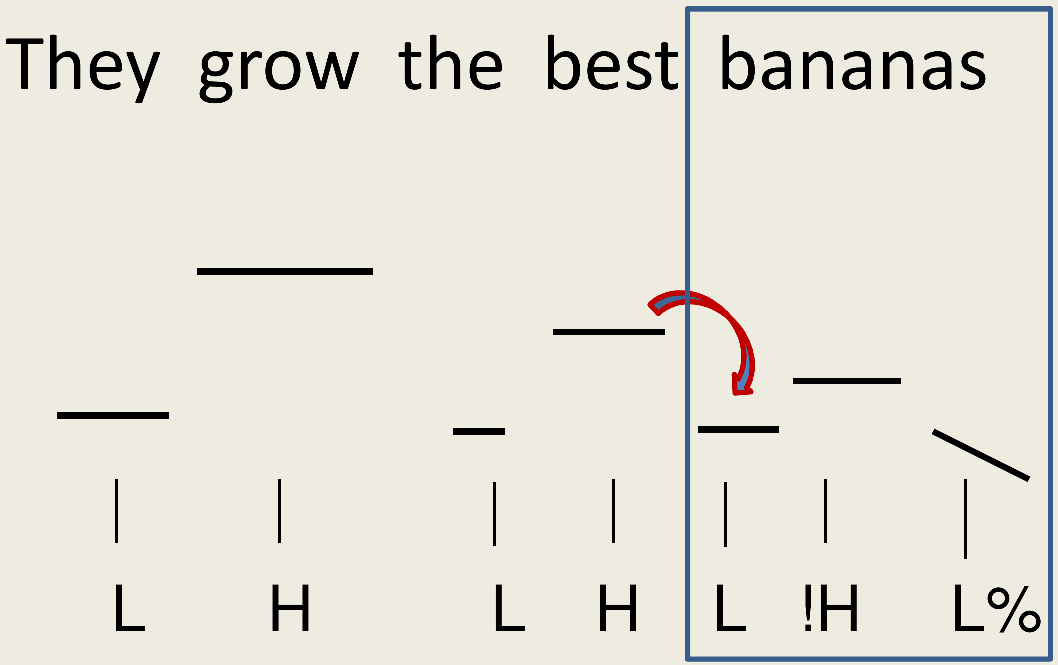
Do lexical words have (L)H or H?

- Check acceptability of:
  - pitch rise to non-initial stress
  - in words like *September, banana*, as compared to words like *Canada, melody*, etc.
  - both IP-finally and IP-medially.
- Check f0 and durations in a production experiment.


They grow the best bananas
L
H
L
H
!H
L%

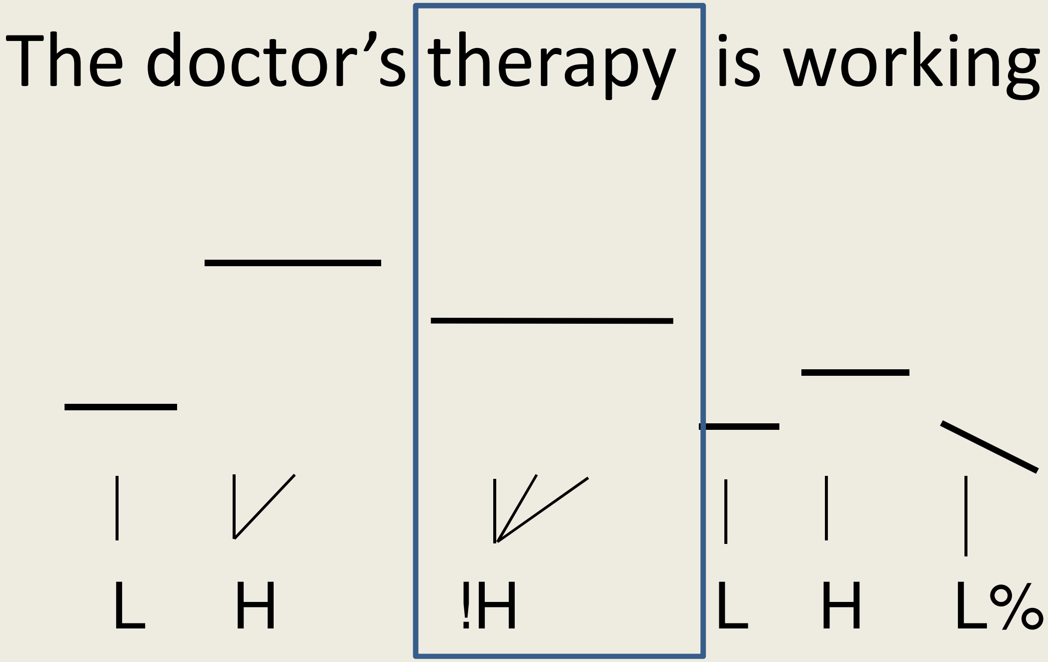
Hypothesis:

- Peninitial stress with H.
- Initial syllable has L.



Hypothesis:

- Initial stress with H.



Acceptability Experiment

(Gussenhoven & Udofot, in prep)

- 6 trisyllabic words embedded in sentences, 3 with penultimate stress (e.g. *employer*), 3 with initial stress (e.g. *cinnamon*).
- 2 speakers (M and F).
- Embedded medially and finally in simplex sentences, always after a lexical word.
- Each of these 24 source utterances was provided with 4 artificial f0 contours.
- 21 SNE listeners (Uyo University)

Four synthetic f0 contours

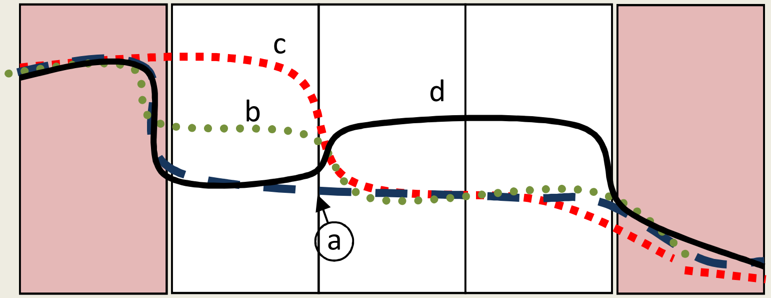
Pitch fall at

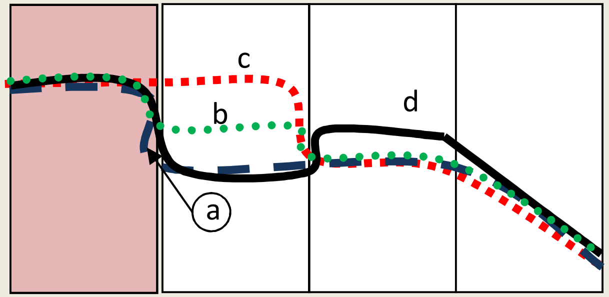
(a) left word boundary.

(b) left word boundary and 2nd syllable boundary.

(c) 2nd syllable boundary.

(d) left word boundary and a *rise* at 2nd syllable.



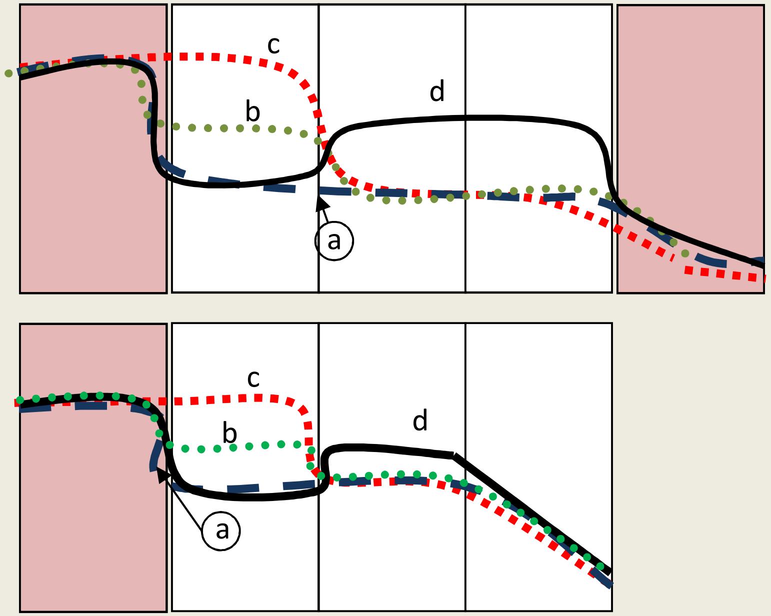


c

Predictions

Initial stress

- fall at word boundary: (a) is best
- level at 2nd syll: (d) is worst

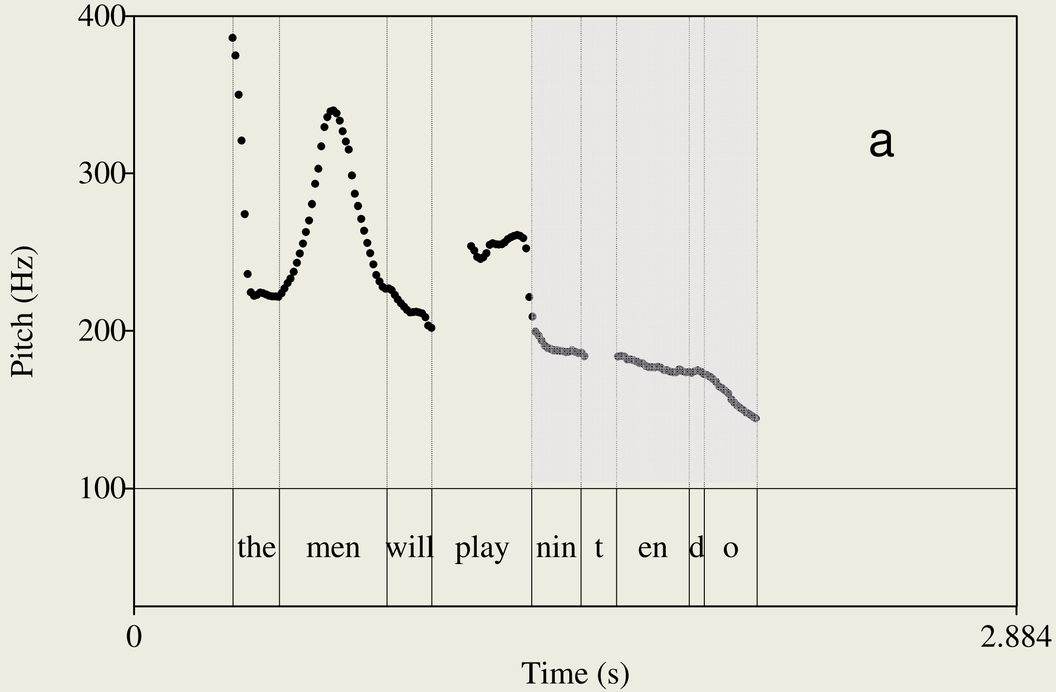


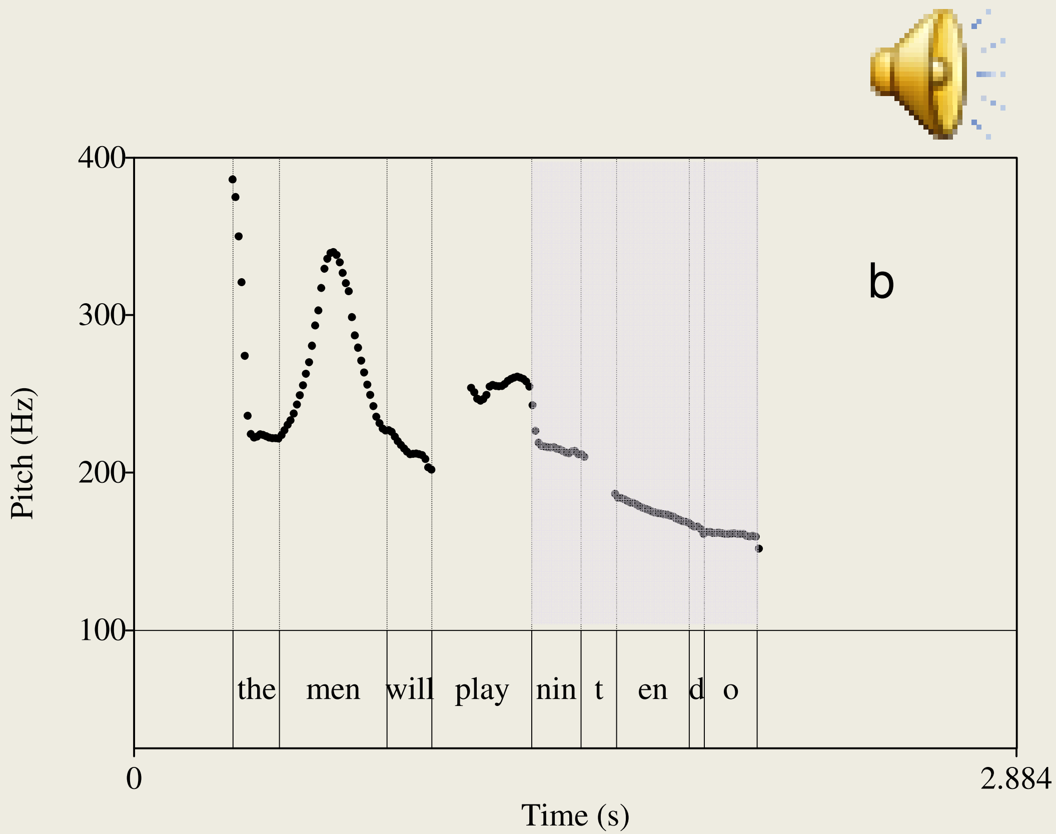
c

Predictions

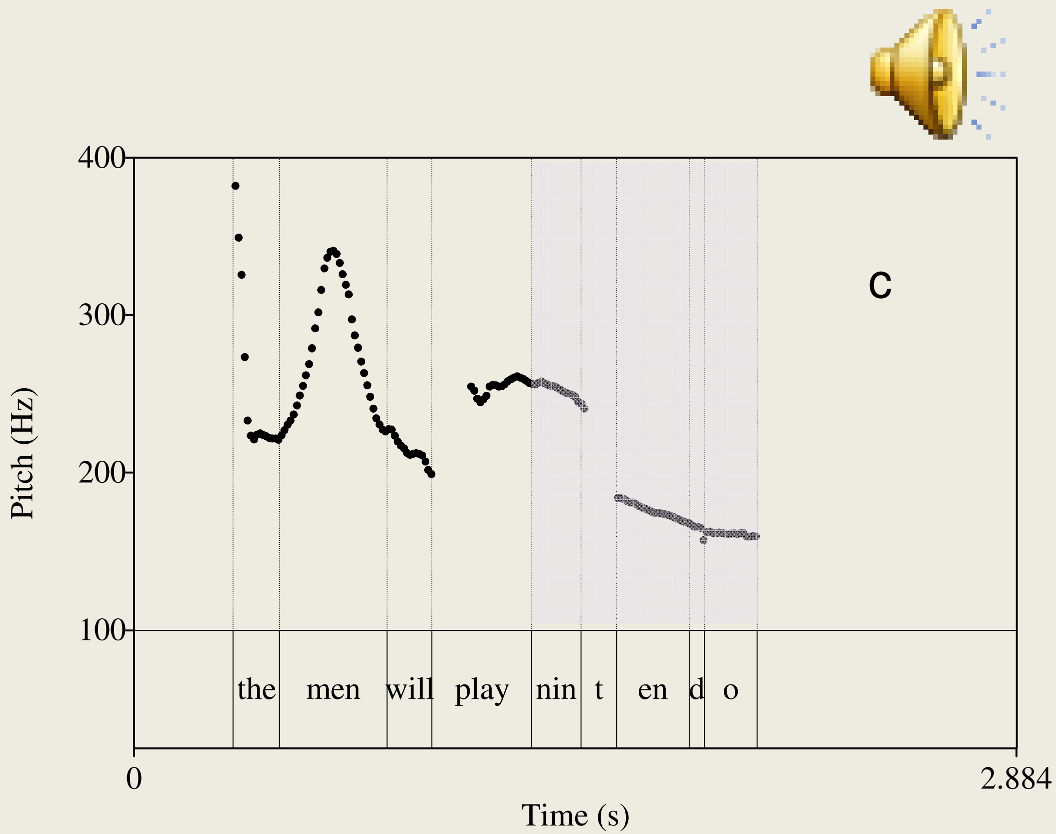
Peninitial stress

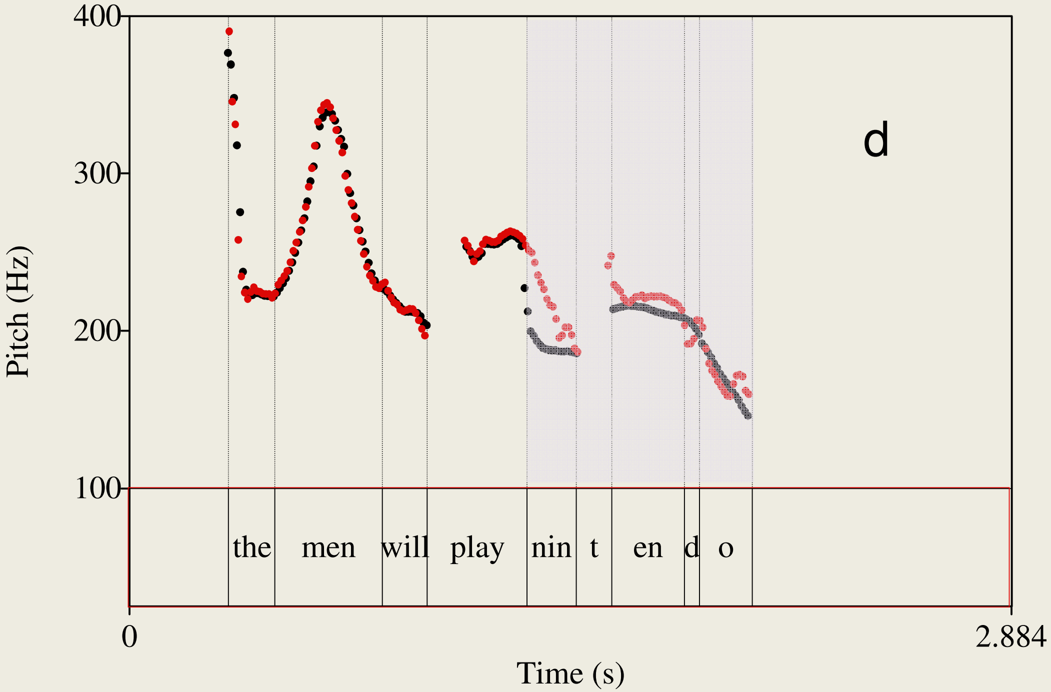
- fall at word boundary, rise at 2nd syll: (d).
- (a)(b)(c) bad.

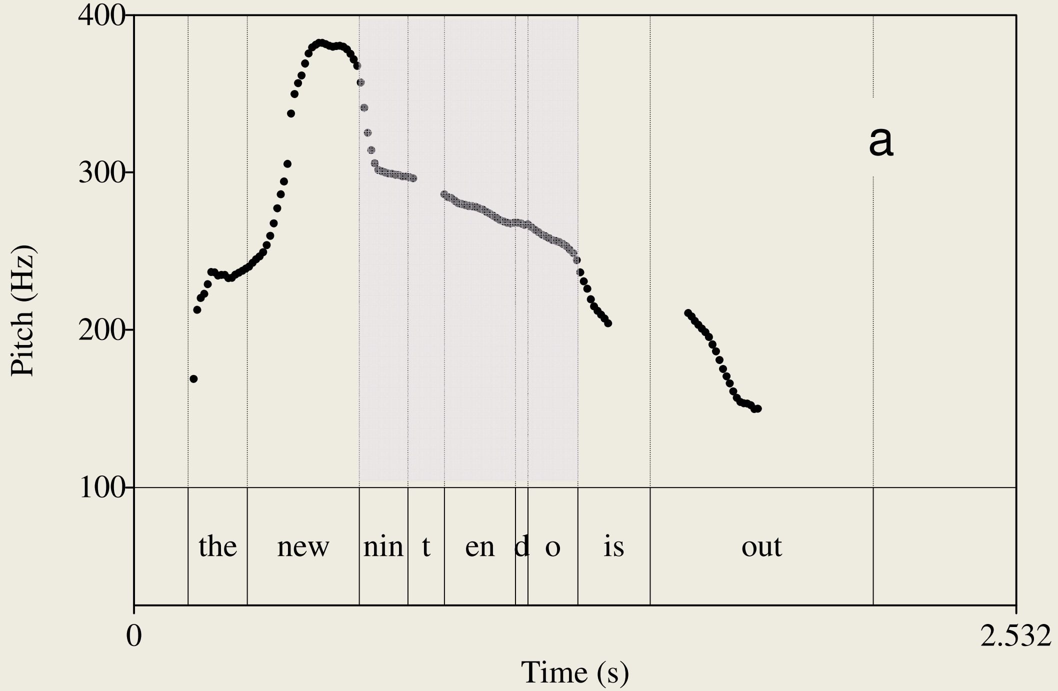

a
400
300
200
100
Pitch (Hz)
the
men
will
play
nin
t
en
d
o
0
2.884
Time (s)

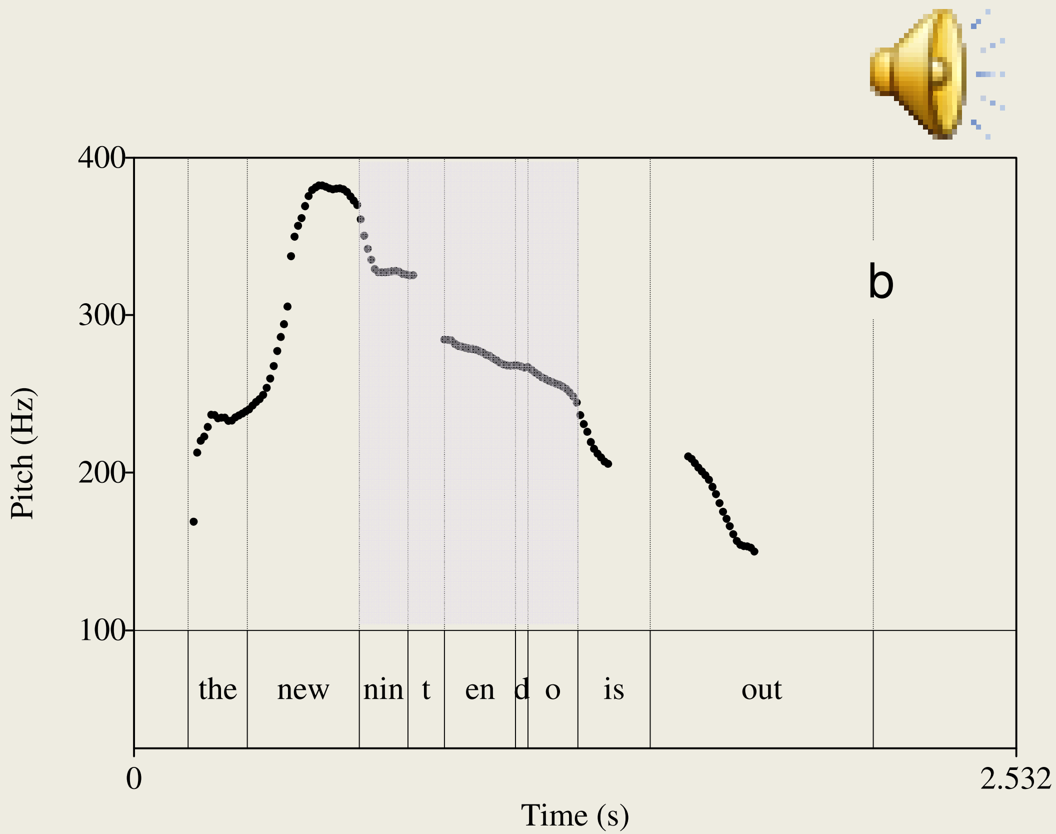


b
400
300
200
100
Pitch (Hz)
the
men
will
play
nin
t
en
d
o
0
2.884
Time (s)

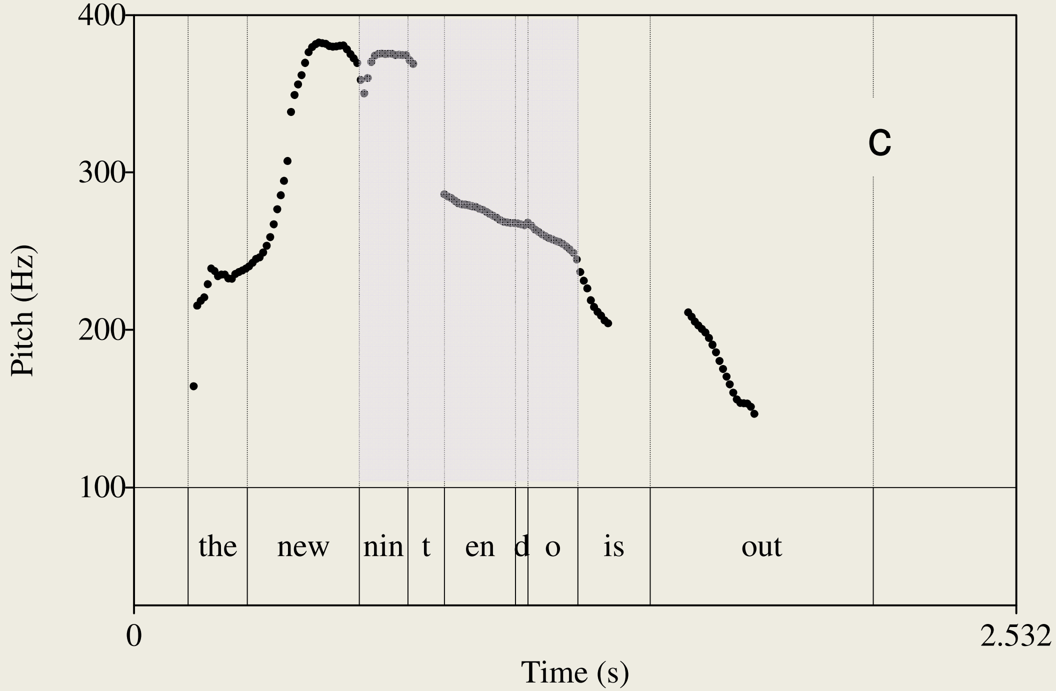

c
400
300
200
100
Pitch (Hz)
the
men
will
play
nin
t
en
d
o
0
2.884
Time (s)

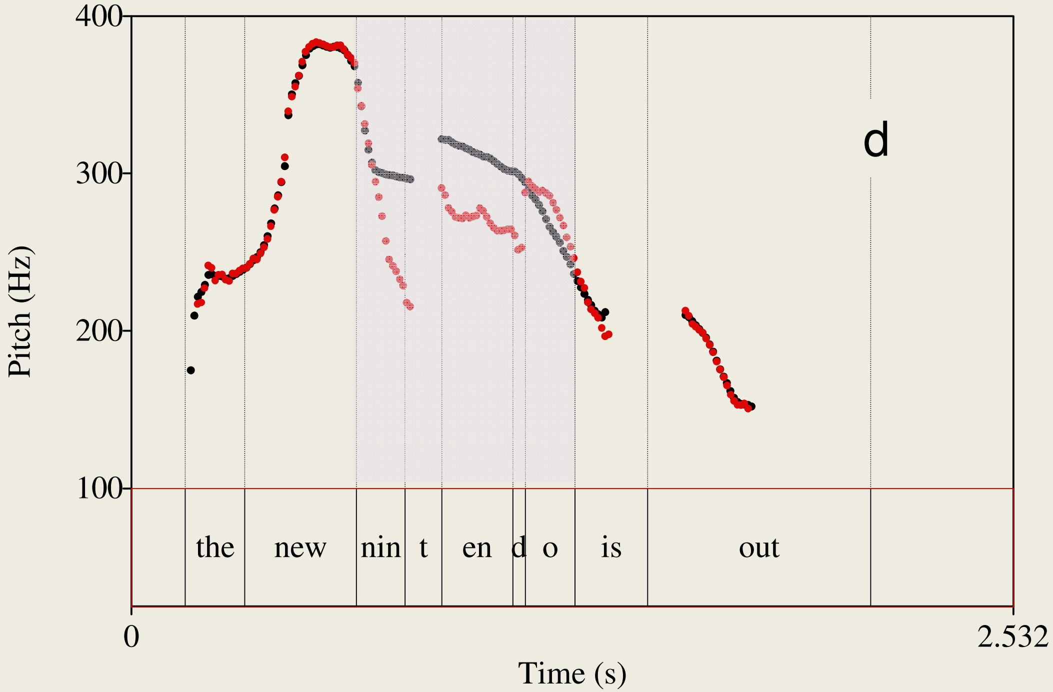

d
400
300
200
100
Pitch (Hz)
the
men
will
play
nin
t
en
d
o
0
2.884
Time (s)


400
300
200
100
Pitch (Hz)
a
the
new
nin
t
en
d
o
is
out
0
2.532
Time (s)

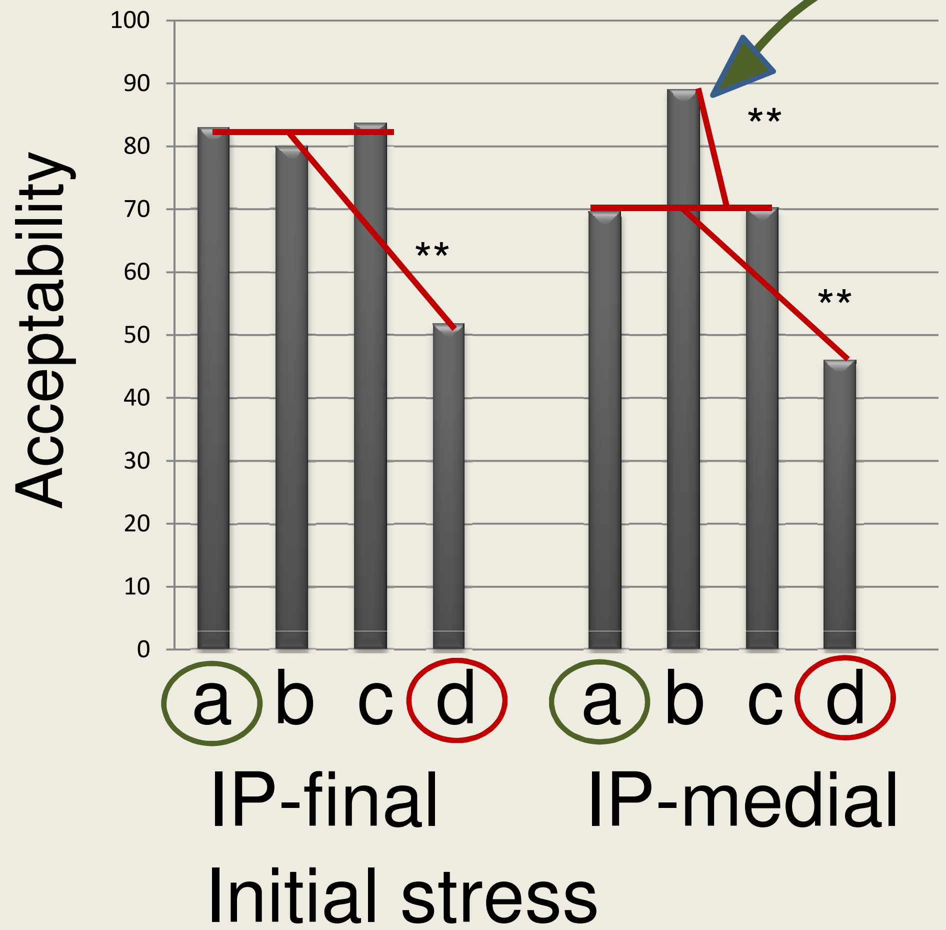


400
300
200
100
Pitch (Hz)
b
the
new
nin
t
en
d
o
is
out
0
2.532
Time (s)


c
400
300
200
100
Pitch (Hz)
the
new
nin
t
en
d
o
is
out
0
2.532
Time (s)

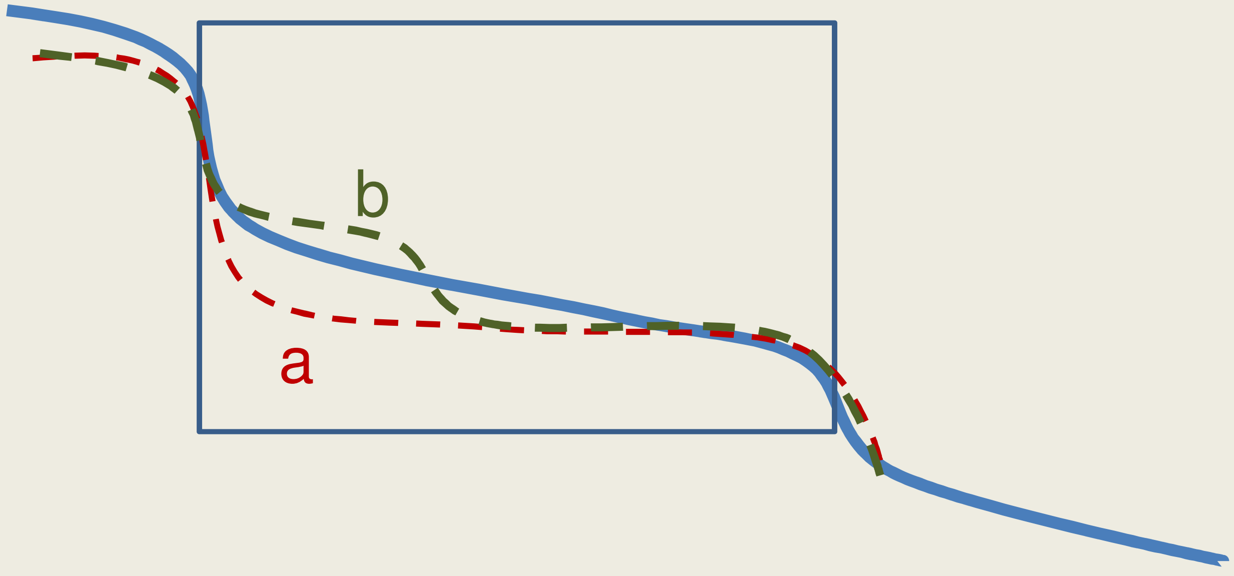


400
300
200
100
Pitch (Hz)
d
the
new
nin
t
en
d
o
is
out
0
2.532
Time (s)



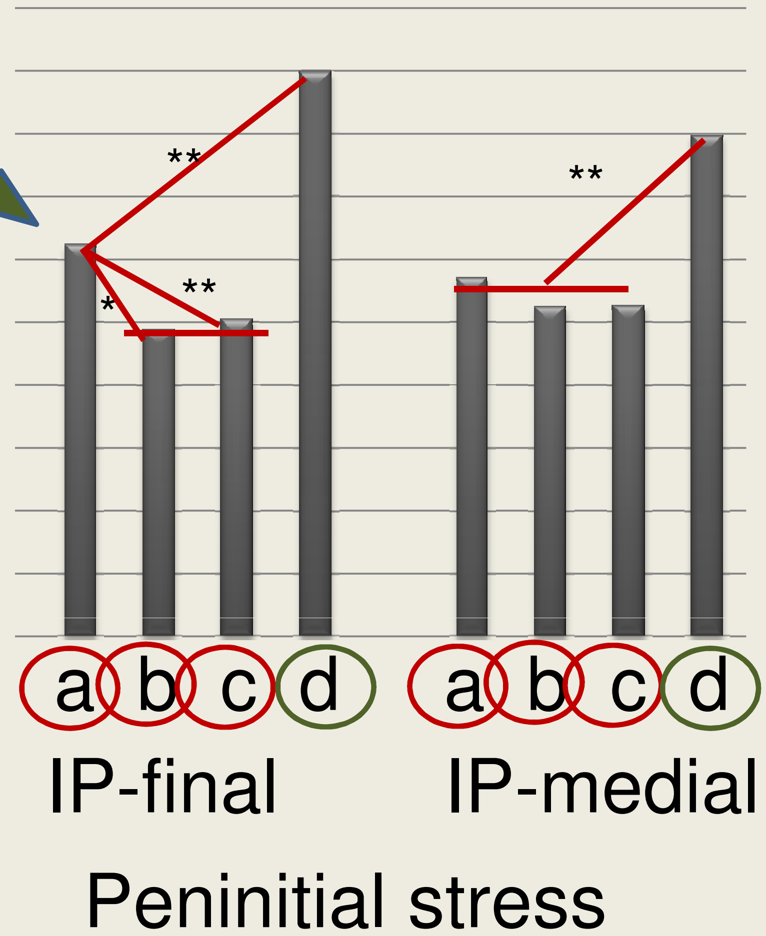
Contour (b) is good, because it better matches the declination between H and !H in the first syllable.

Medial: wider pitch range.


b
a

Contour (a) is relatively good, because it *doesn't fall* to stress where it should rise. (b) and (c) *fall* to stress.

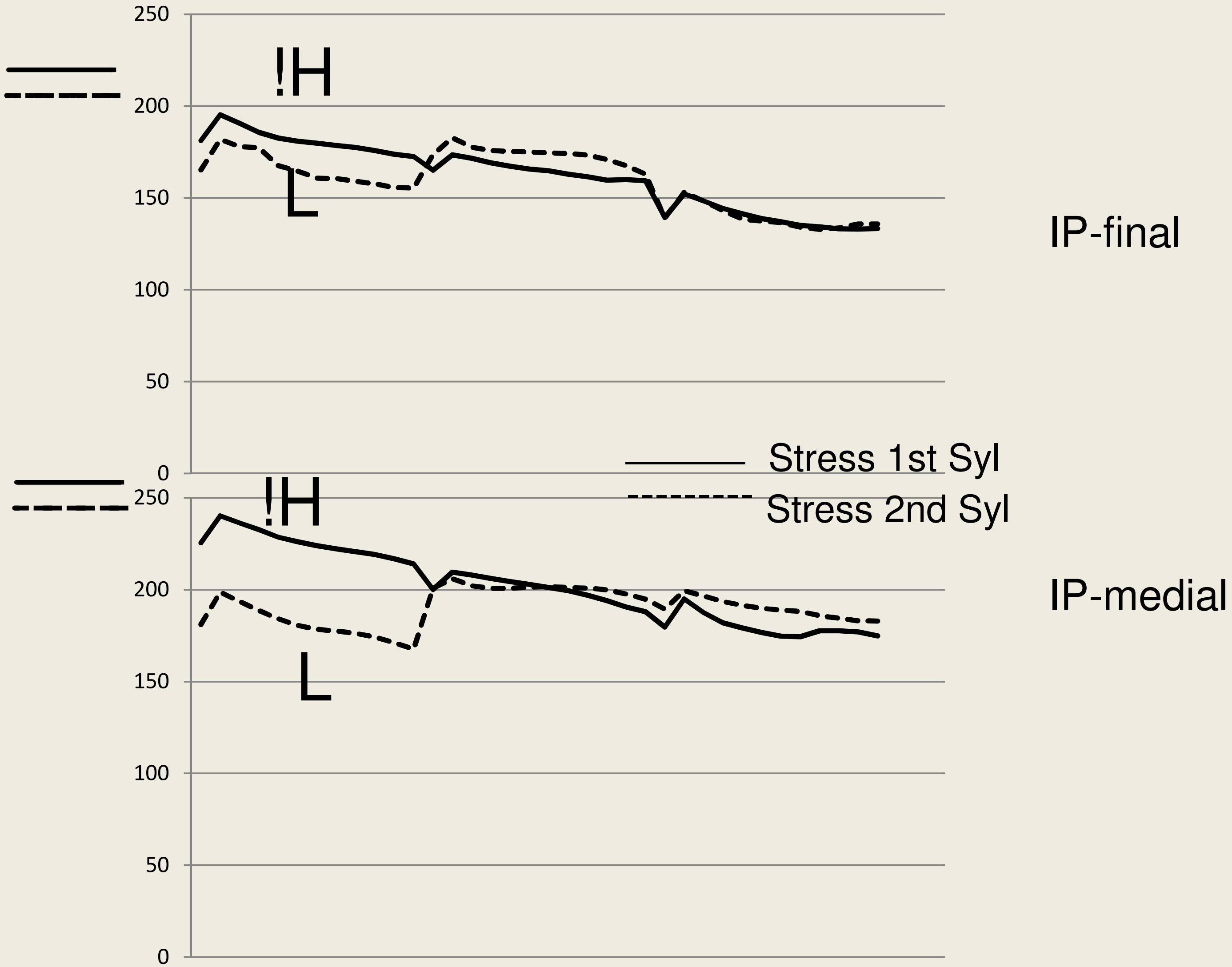
Final: more salient.



Production Experiment

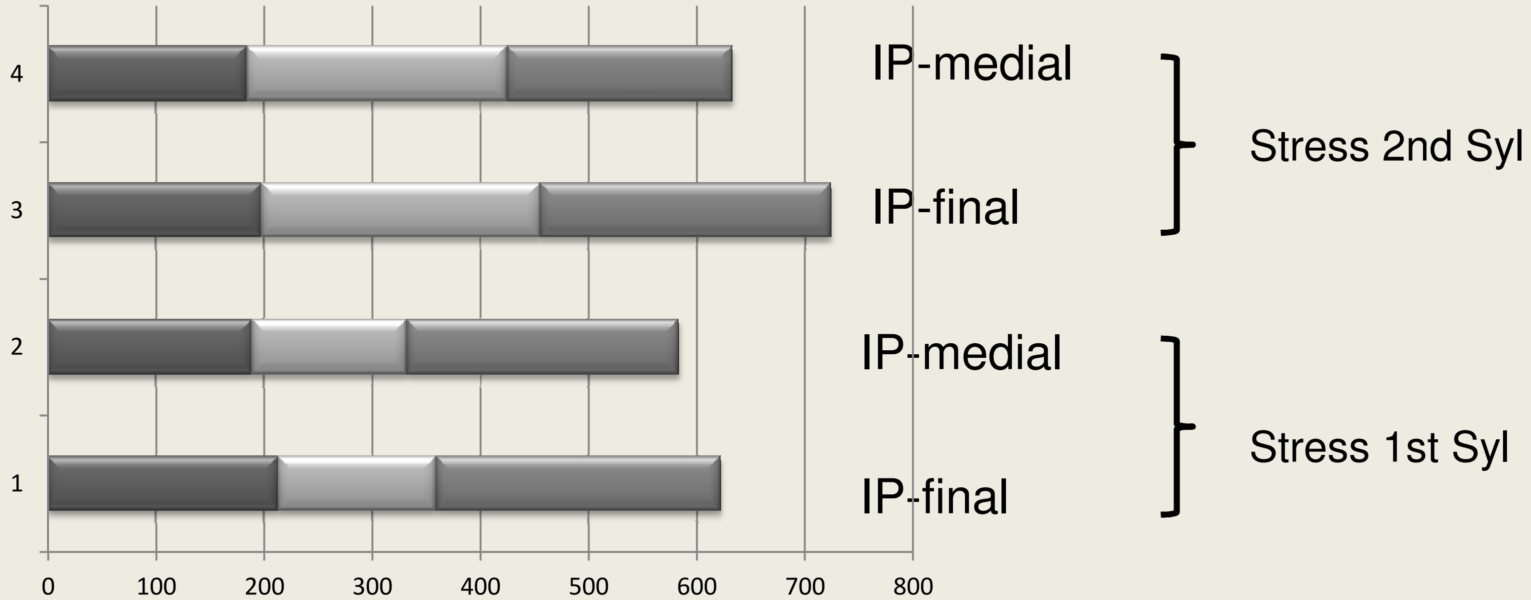
- 1 male, 3 female speakers of SNE (SOAS).
- 18 trisyllabic words embedded in sentences, 9 with penultimate stress (e.g. *September, employer*), 9 with initial stress (e.g. *cinnamon, spectacle*).
- Embedded medially and finally in simplex sentences.
- Measurement of f0 and duration of all syllables

Results f0

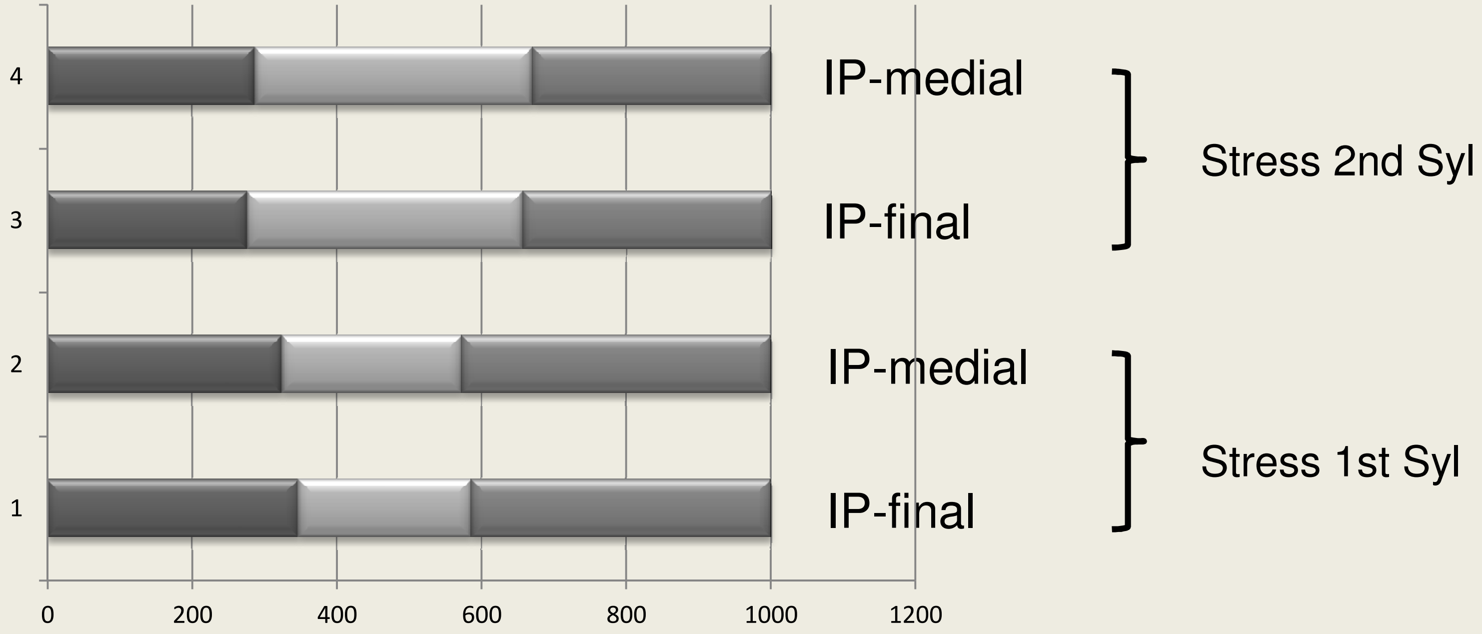


250
200
150
100
50
0
!H
L
IP-final
Stress 1st Syl
Stress 2nd Syl
250
200
150
100
50
0
!H
L
IP-medial

Results duration



Results duration



Conclusion for Nigerian English

1. It has (initial vs peninitial) stress.
2. It has tone: lexical words have LH, with H going to the first or second syllable.
3. L causes downstep (whether floating or not).
3. Function words have L.

What's so good about Nigerian English?

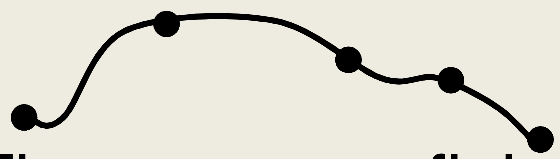
- **Contrast between compound and single word**

  syntax – sin tax; insect - inn sekt

- **Contrast between function words and lexical words**

  would – wood; in – inn; are – R; were – whirr, etc.

Minimal pairs

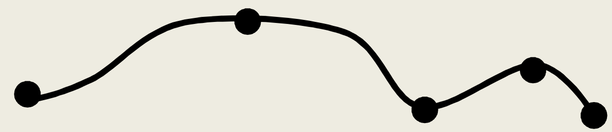


The men can fish

| | | |

L *L*H *L*H *L*HL%

The men put fish in tins



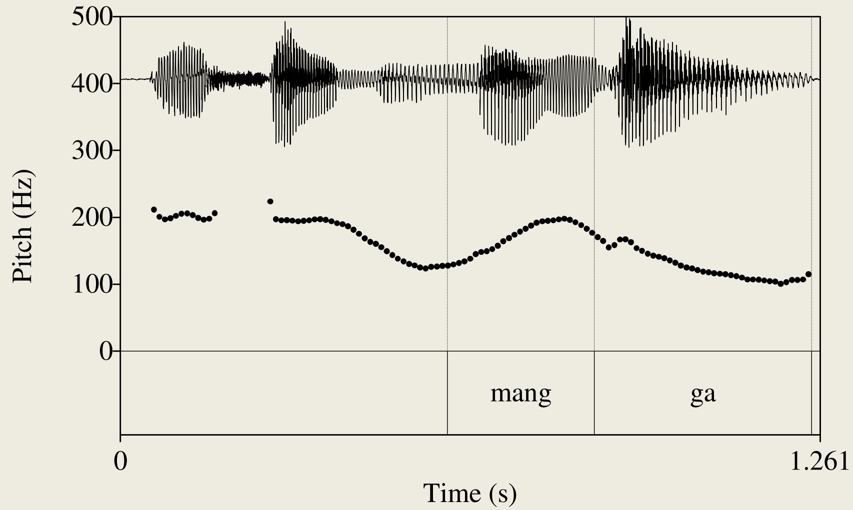
The men can fish

| | | |

L *L*H L *L*HL%

The men are able to fish

Ambonese Malay:
No stress, no pitch accent, no focus



Dorang suka makang pisang deng mangga

'They would like to eat a banana and a mango.'

Possible intonational tone structures for a final peak

(1) a. … ma ma **ma** $)_{\omega})_{\varphi})_{\iota}$

|

H*+L

b. … ma ma ma $)_{\omega})_{\varphi})_{\iota}$

|

H* L%

c. … ma ma ma $)_{\omega})_{\varphi})_{\iota}$

|

H% L%

d. … ma ma ma $)_{\omega})_{\varphi})_{\iota}$

H% L%

Two arguments for stress in Ambonese Malay:

1. Stress location is rule governed.

2. There are minimal pairs (Ambonese Malay).

Variation among 13 authors

‘Stress is penultimate’
Two sources of variation:

- schwa in the penultimate syllable
- intonational melody.

-Penultimate schwa moves stress to final syllable in disyllables, but to antepenult in trisyllabic or longer words, unless the penult also has schwa (Halim 1974).
-Stress is penultimate without exception, i.e. also when it has schwa (Laksman 1974) .
- Penultimate schwa moves stress to final syllable (Java), unless 2 consonants separate schwa from final syllable (outside Java) (Prentice 1987)..
- In pre-final pause groups, stress is final, in final pause groups, stress is penultimate (Halim 1974).

(Odé 1994)

Minimal pairs

Beta mobilan	parang	*'machete'*
'I say'	parang	*'war'*
	masing	*'salty'*
	masing	*'machine'*
	barat	*'West'*
	barat	*'heavy'*

van Minde 1997

Minimal pairs

But /a/ would be the only vowel that can reject stress. There is no penultimate /i, e, o, u/ without stress.

Short /a/ is like schwa in Indonesian.
It's an a-*caduc*.
There are six vowels, not five.

French

Phonetic signal without stress

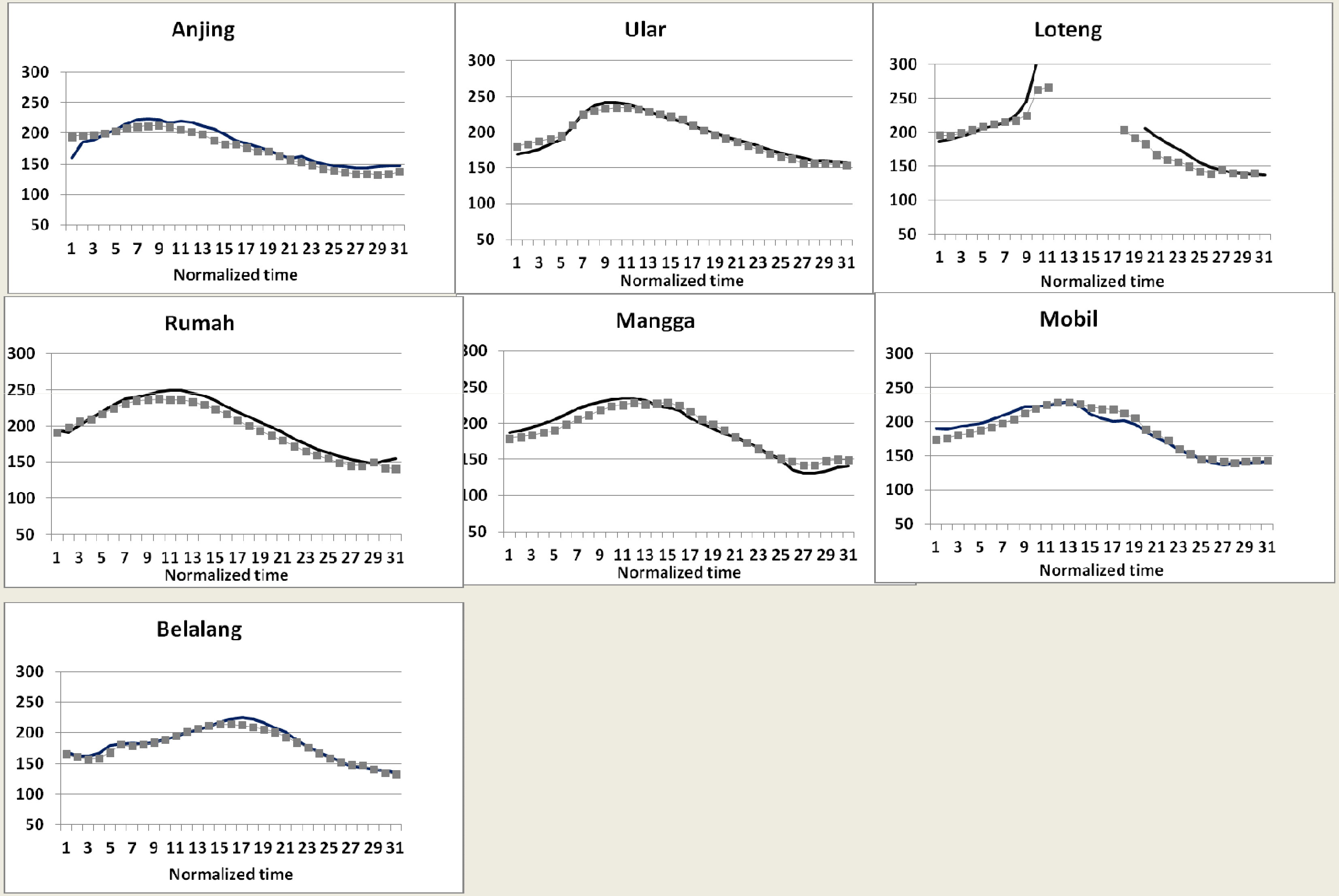
French [pəle] *peler* ‘peal’
[pele] *Pelé* ‘Edson Arantes do Nascimento’
(e-*caduc*)

Peak alignment

Raechel Maskikit & C. Gussenhoven, submitted

- 2 female and 2 male speakers from Ambon.
- Six disyllabic target words, 1 trisyllabic.
- Declarative/Interrogative/Prefinal.
- Repeated word/Corrective focus.

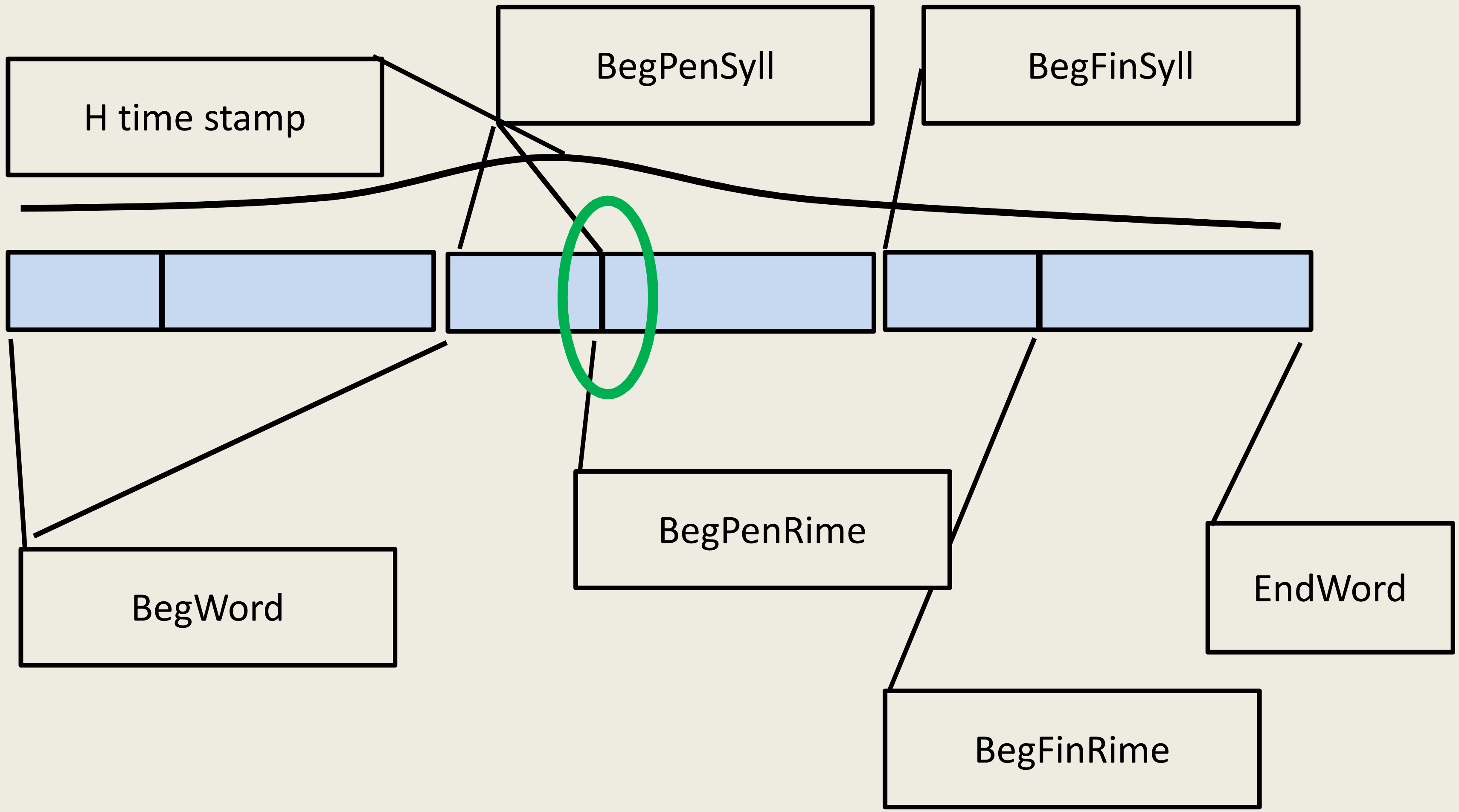
No difference between Given and Corrective focus



Matched Ambonese Malay and Dutch corpora

anjing	‘dog’	**An**jum	
balalang	‘grasshopper’	be**ma**ling	‘drainage’
loteng	‘attic’	**lo**ting	‘lottery’
mangga	‘mango’	**man**go	
mobil	‘car’	**meu**bel	‘piece of furniture’
rumah	‘house’	**Ro**ma	
ular	‘snake’	**Oe**ral	

F0: the alignment of peak



Alignment: Two types of analysis

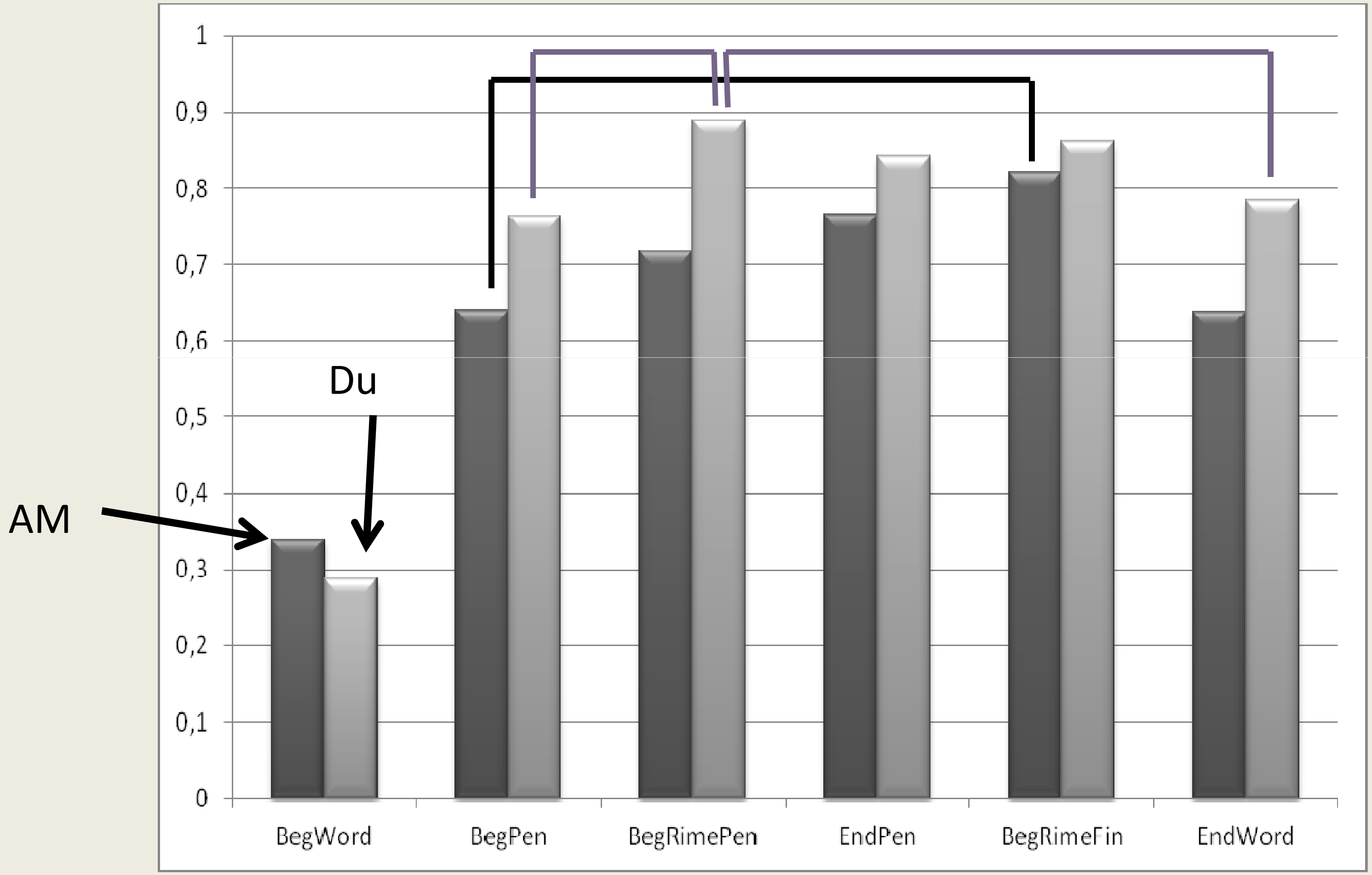
- Time stamp analysis:

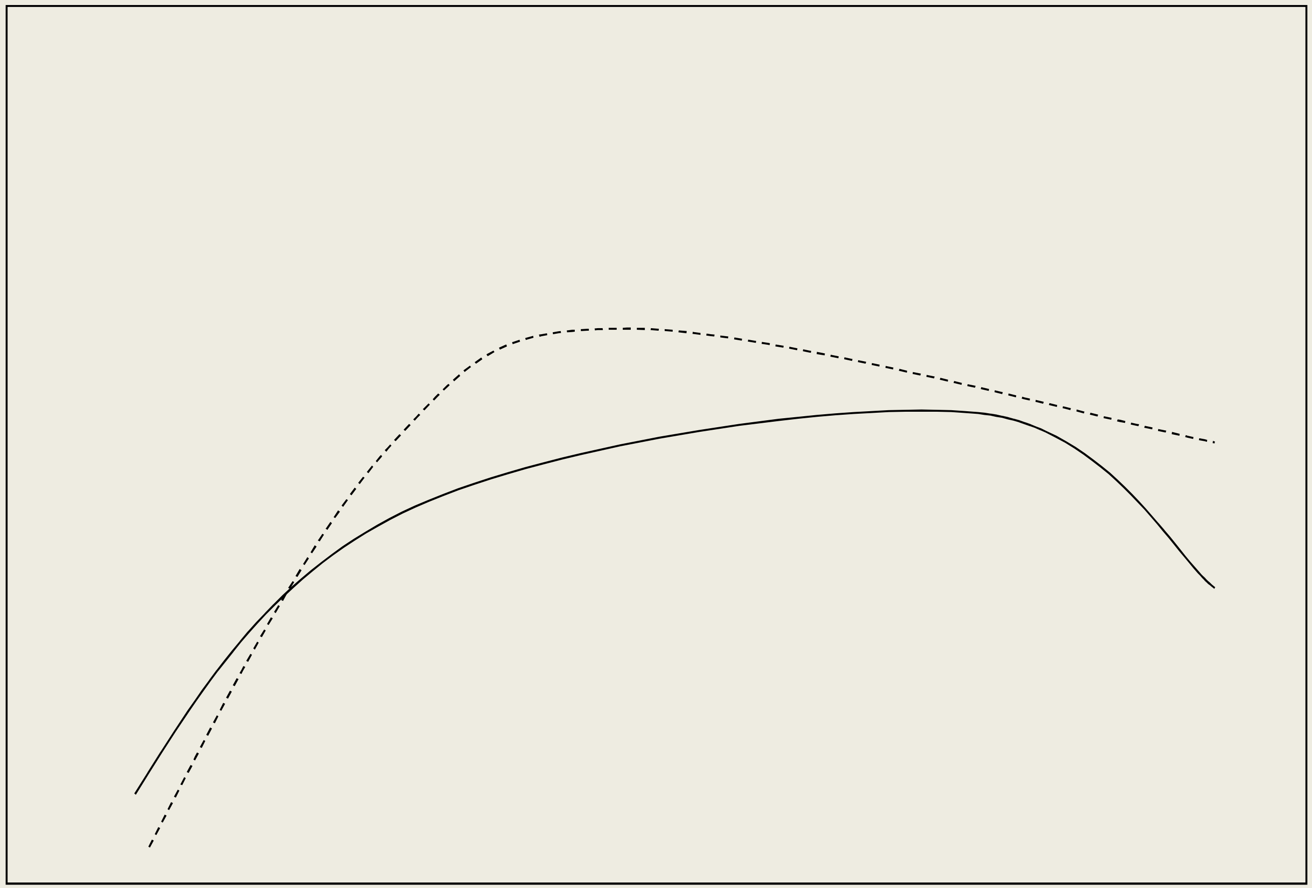
Does time stamp of H correlate with time stamp of landmark?

- Latency analysis:

Does the distance between H and landmark covary with duration of rime/syllable/word?

Pearson r's between H time stamp and time stamps of six landmarks (N=84)





Welby (2006)

Regression with H-latency from beginning of final syllable as dependent variable and duration of final syllable:

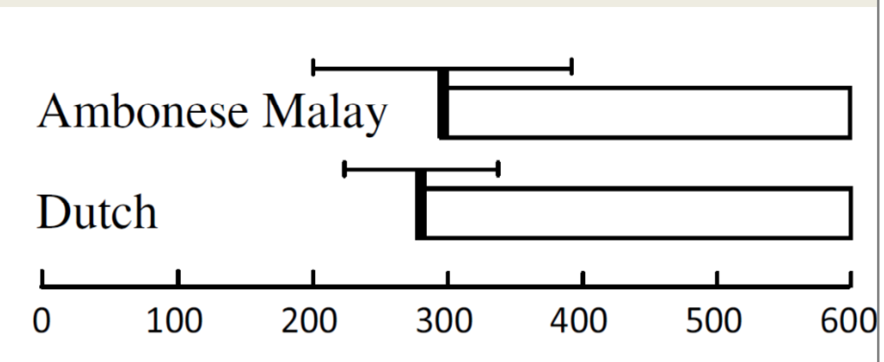
Together with a binary variable (LHLH vs Other), 74% explained variance.

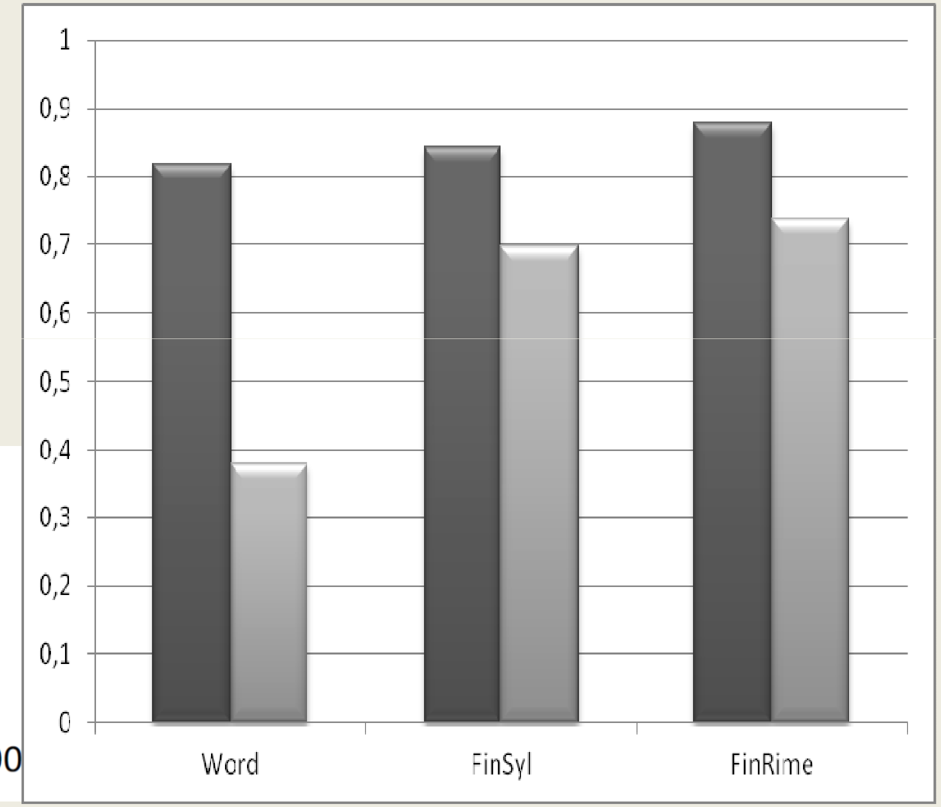
Highly regular location, calculated from syllable beginning as a function of syllable duration.

H-location: 60 ms before start of syllable, with L following10 ms after start of last syllable.

Regression with H-latency from beginning of final syllable as dependent variable and final syllable duration: 3%, with word duration added: 30%.

Pearson r's between **H-Latency to word end** and duration of final constituent.



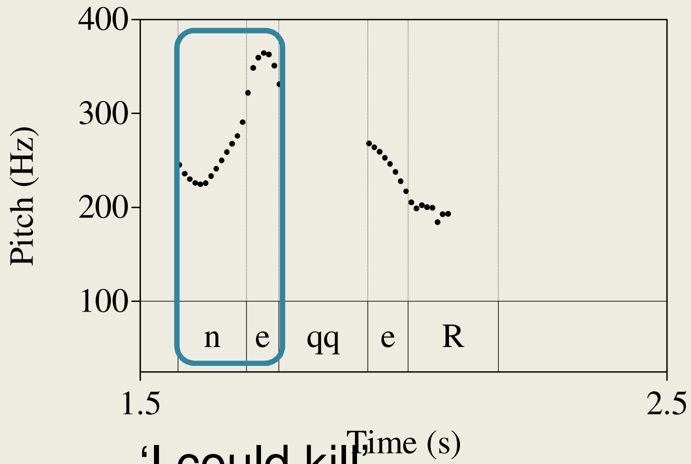


Conclusion Ambonese Malay

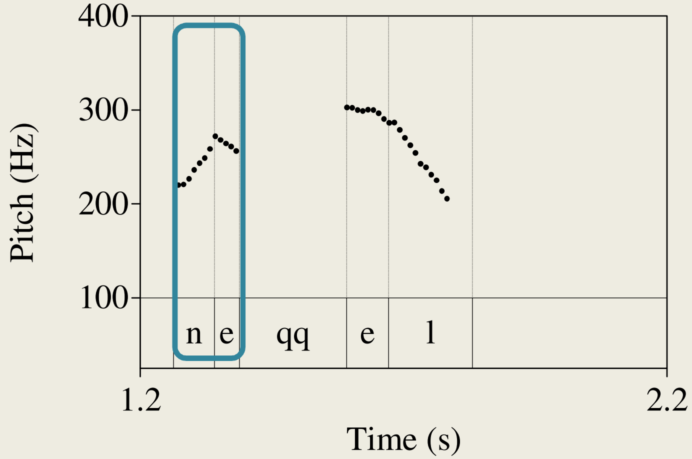
- Ambonese Malay places the pitch peak within the last word, with considerable variation in segmental alignment, relating it to the word end.
- Dutch aligns the pitch peak with the beginning of the stressed rime.
- Ambonese Malay has no stress, no pitch accent, no prosodic focus.
- It has a-*caduc.*

Zwara Tamazight (Libya)

Gussenhoven in prep. (TOPIQQ, Dobes)



‘I could kill’



‘we agreed’

Zwara segments

	Labial	Alveolar	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	b	t ṭ d ḍ		k g	q (q')		
Nasal	m ṃ	n ṇ					
Fricative	f	s ṣ z ẓ	ʃ ʒ		χ (χ') ʁ (ʁ̣)	ħ ʕ	ɦ
Approximant	w	r ṛ	j				
Lateral		l ḷ					

i ɪ/ə æ u

ɨ ɘ/ɜ ɑ o

(C) V (C)(C)

j V j C j C

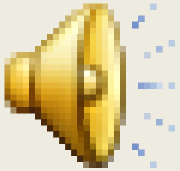
i V i C i C

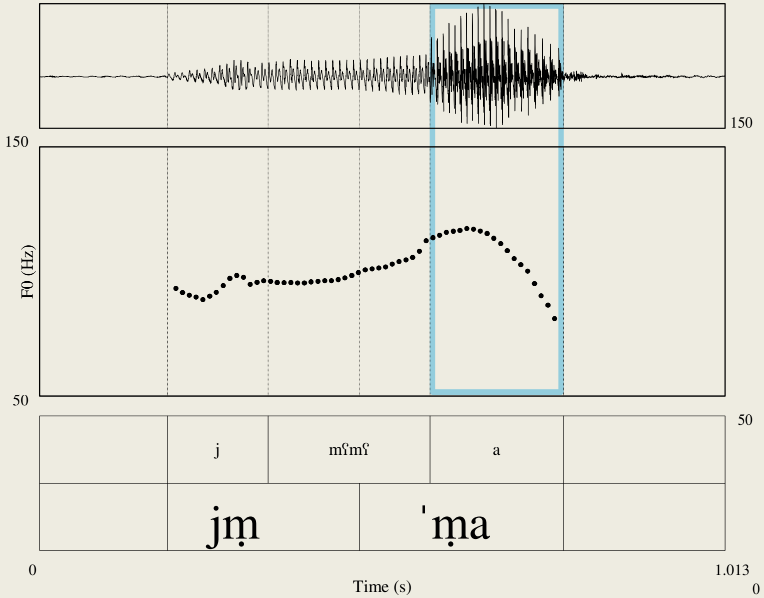
No hiatus: VV VCV

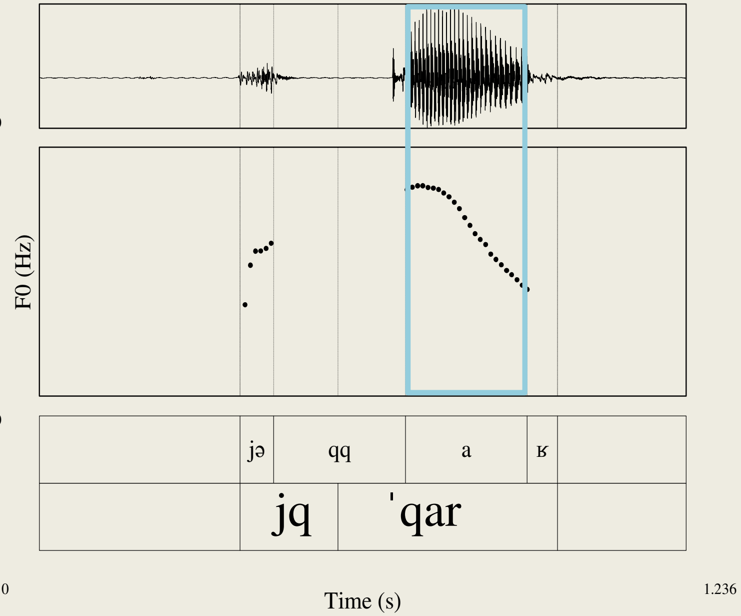
Zwara stress

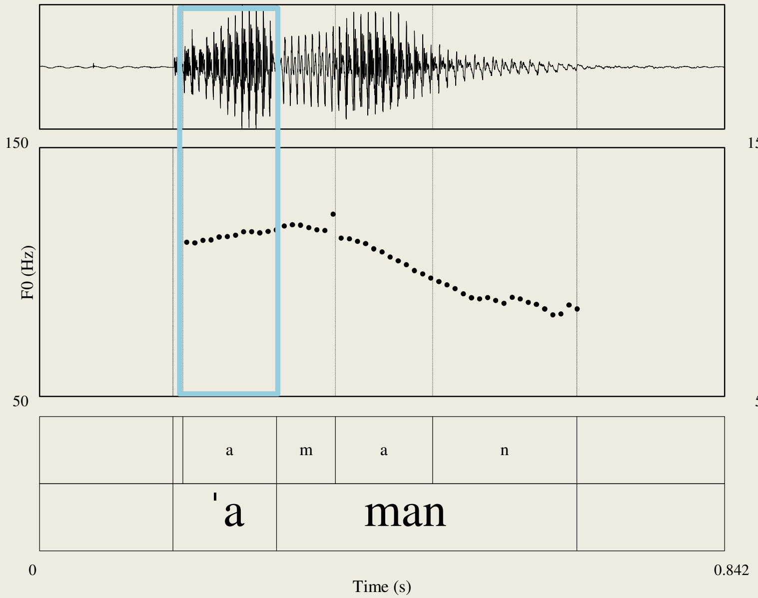
1. Penultimate
2. Final, e.g. in PAST verbs.
3. Pre-stressing negative and question suffixes.

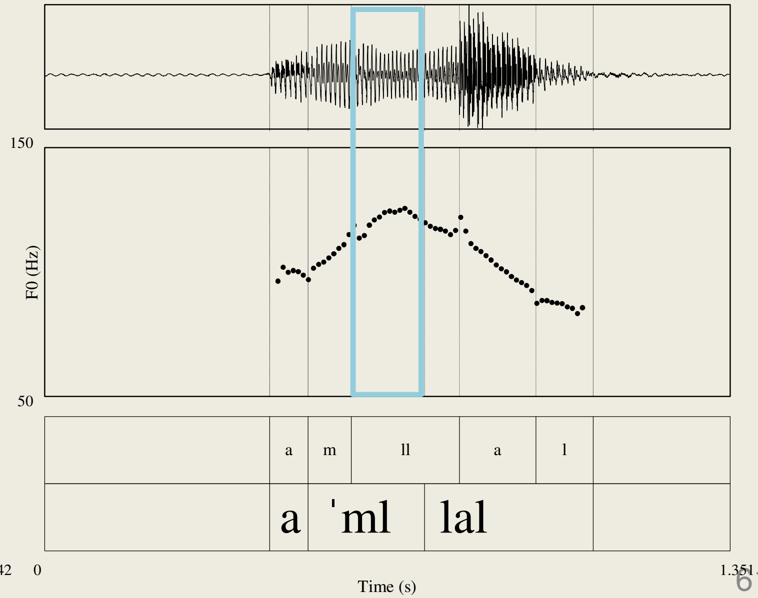
jṃ.ˈṃa	‘he said’
jq.ˈqar	‘he was reading’
ˈa.man	‘water’
a.ˈml.lal	‘white’
ˈə.wtt	‘hit-IMP’




150
50
F0 (Hz)
j
mˤmˤ
a
jṃ
ˈṃa
0
1.013
Time (s)

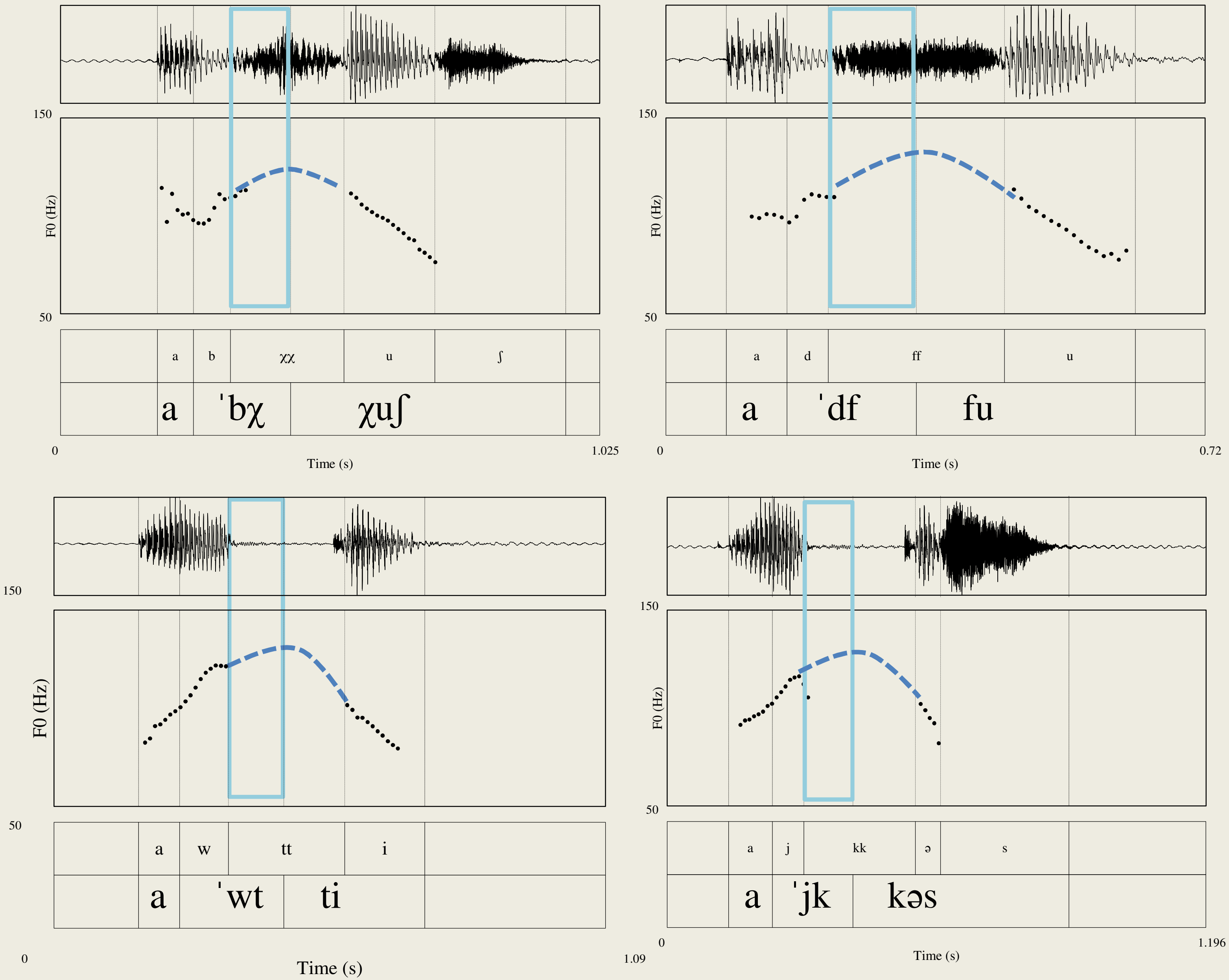

150
50
F0 (Hz)
jə
qq
a
ʁ
jq
ˈqar
0
1.236
Time (s)


150
50
F0 (Hz)
a
m
a
n
ˈa
man
0
0.842
Time (s)


150
50
F0 (Hz)
a
m
ll
a
l
a
ˈml
lal
0
1.351
Time (s)

Zwara stress

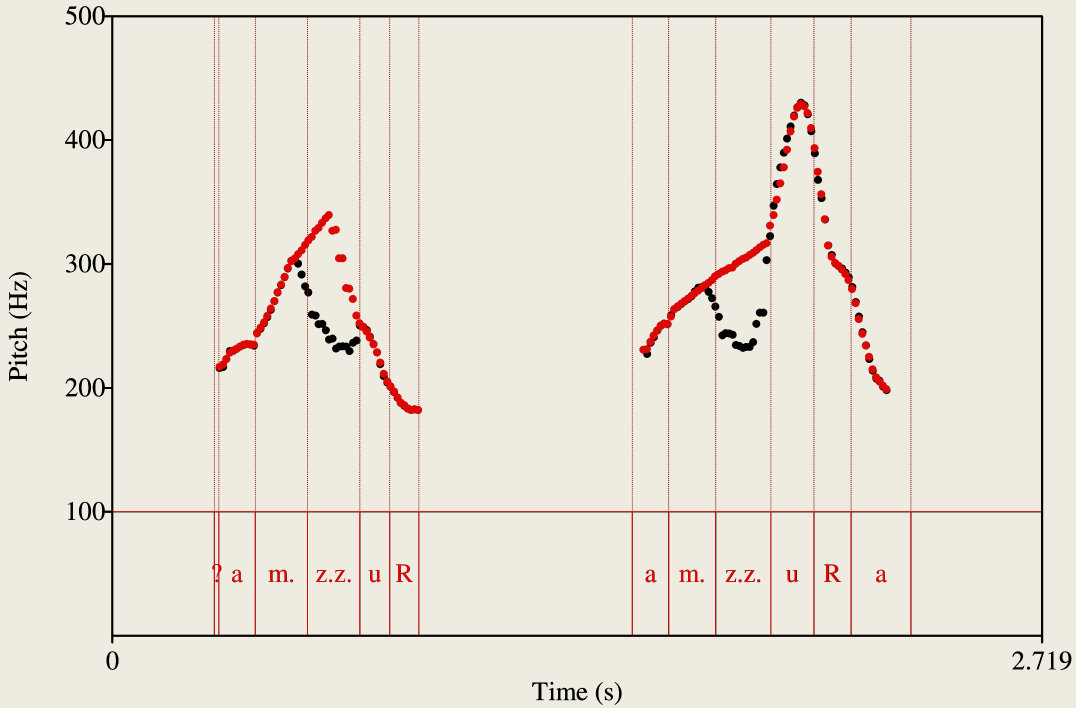
a.ˈbχ.χuʃ	‘insect’
a.ˈdf.fu	‘apple’
a.ˈwt.ti	‘they are preparing’
a.ˈjk.kəs	‘he will take off’


150
50
F0 (Hz)
a
b
χχ
u
ʃ
a
ˈbχ
χuʃ
0
1.025
Time (s)
150
50
F0 (Hz)
a
d
ff
u
a
ˈdf
fu
0
0.72
Time (s)
150
50
F0 (Hz)
a
w
tt
i
a
ˈwt
ti
0
1.09
Time (s)
150
50
F0 (Hz)
a
j
kk
ə
s
a
ˈjk
kəs
0
1.196
Time (s)

[[....] a]   Polar question particle

a.ˈbl.bul	a.bl.ˈbu.la	‘An octopus?’
a.ˈdf.fu	a.df.ˈfu.a	‘An apple?’
ṭa.ˈẓu.ḍa	ṭa.ẓu.ˈḍa.a	‘A plate?’
jṃ.ˈṃa	jṃ.ˈṃa.a	‘He said?’
ˈjm.ma	jm.ˈma.a	‘My mother?’

F0: It's the thought that counts

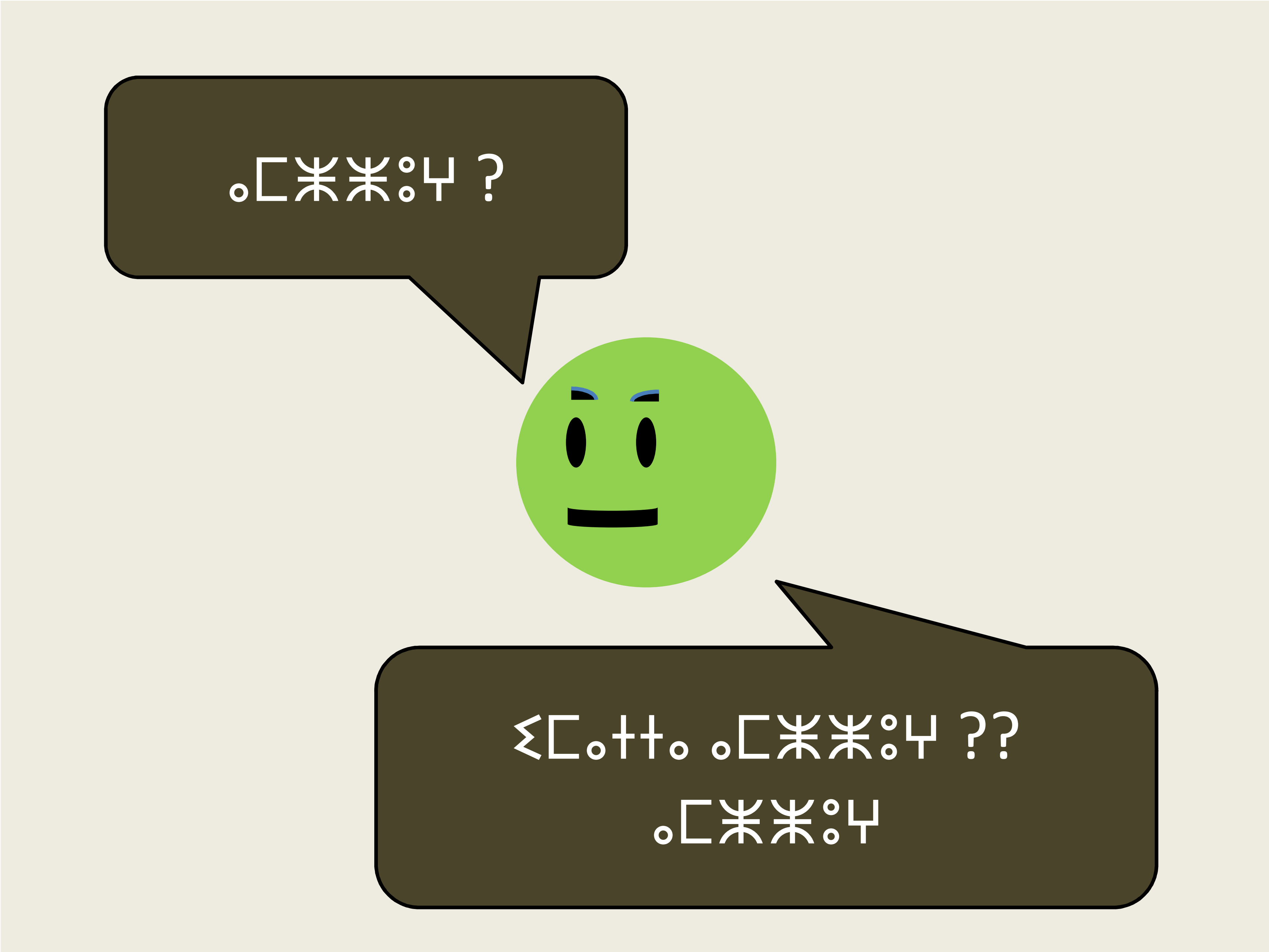


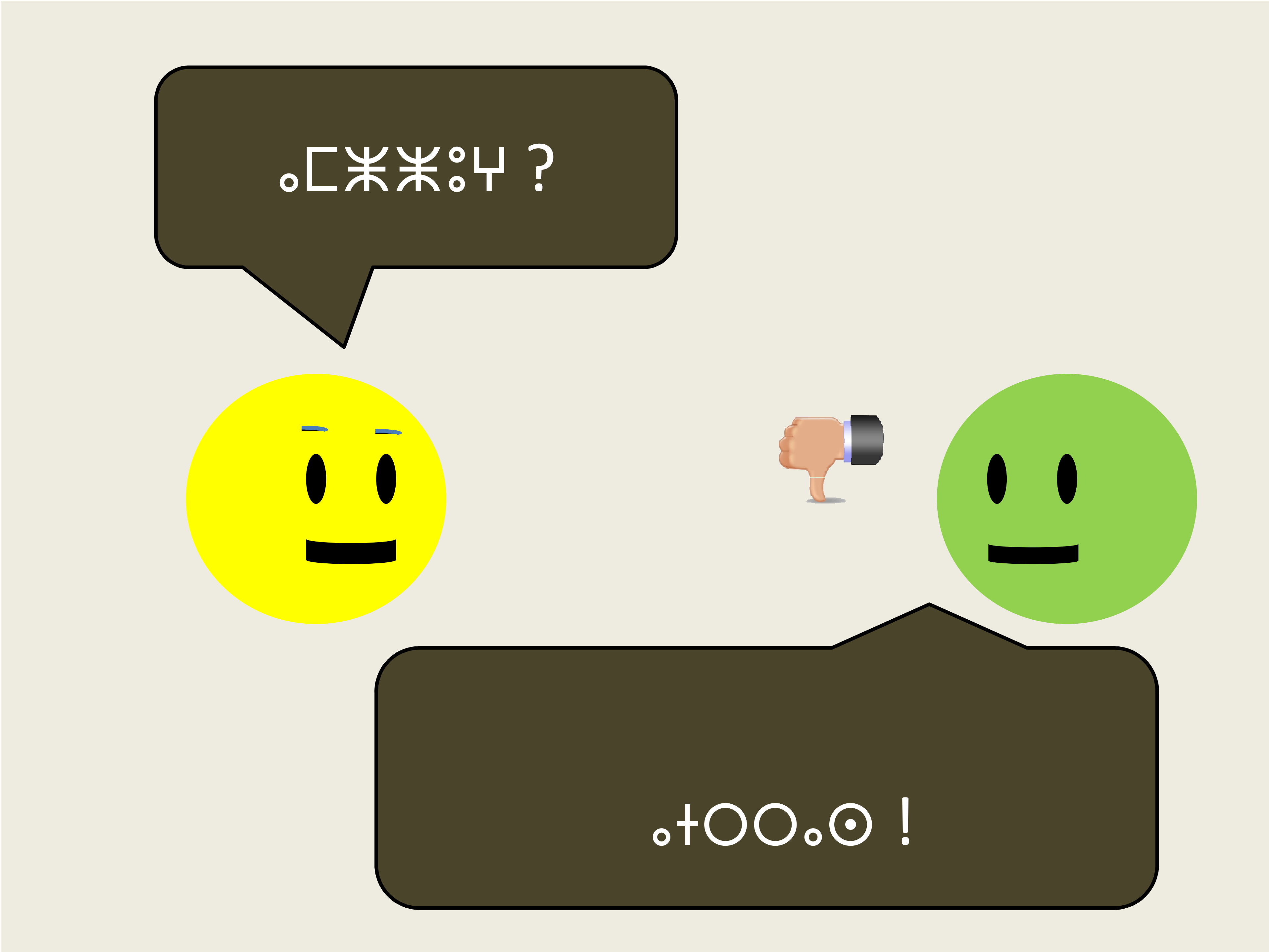
Location for pitch accents

Say X? Why should I say X? I just said X

/jXs/ ‘he wants to’

jXs	**HLH**	**LH**	**HL**


ⴰⵎⵥⵥⵓⵖ ?
ⵉⵎⴰⵜⵜⴰ ⴰⵎⵥⵥⵓⵖ ??
ⴰⵎⵥⵥⵓⵖ


ⴰⵛⵥⵥⵓⵖ ?
ⴰⵜⵔⵔⴰⵙ !

Location for pitch accents

Did I say X? No! (I didn't say X!) I said Y.

/jm.'mʕa/ 'he said'
/'jm.ma/ 'my mother'

HLH	HL	HL
'jm.ma		jm.'mʕa
jm.'mʕa	jm.'mʕa	'jm.ma


Conclusion Zuara Tamazight

- Penultimate and final stress.
- Consonants and vowels can be the nucleus of a syllable.
- Stress can occur on silent closure phase of voiceless plosive.
- Stress is location for intonational pitch accent.

Persian word accent

(Abolhasanizadeh, Bijankhan & Gussenhoven 2011)

- Persian words have H-tone on final syllable.
  Ignore clitics: *tób-esh-e* ‘it is his swing’
  tobésh-e ‘it is light’.
- Morphological accent rules.
- No phonetic stress.
- Are speakers of Persian ‘stress deaf’?

(Peperkamp & Dupoux 2002 and following work)

Testing Persian word accent in stress deafness paradigm

(Rahmani, Rietveld & Gussenhoven, in prep.)

- Lower stress deafness baselines: *French*, *Indonesian*.
- Upper stress deafness baselines: *Dutch*, *Japanese*.
- Disyllabic stimuli for ‘stress’ and ‘phoneme’ distinctions:

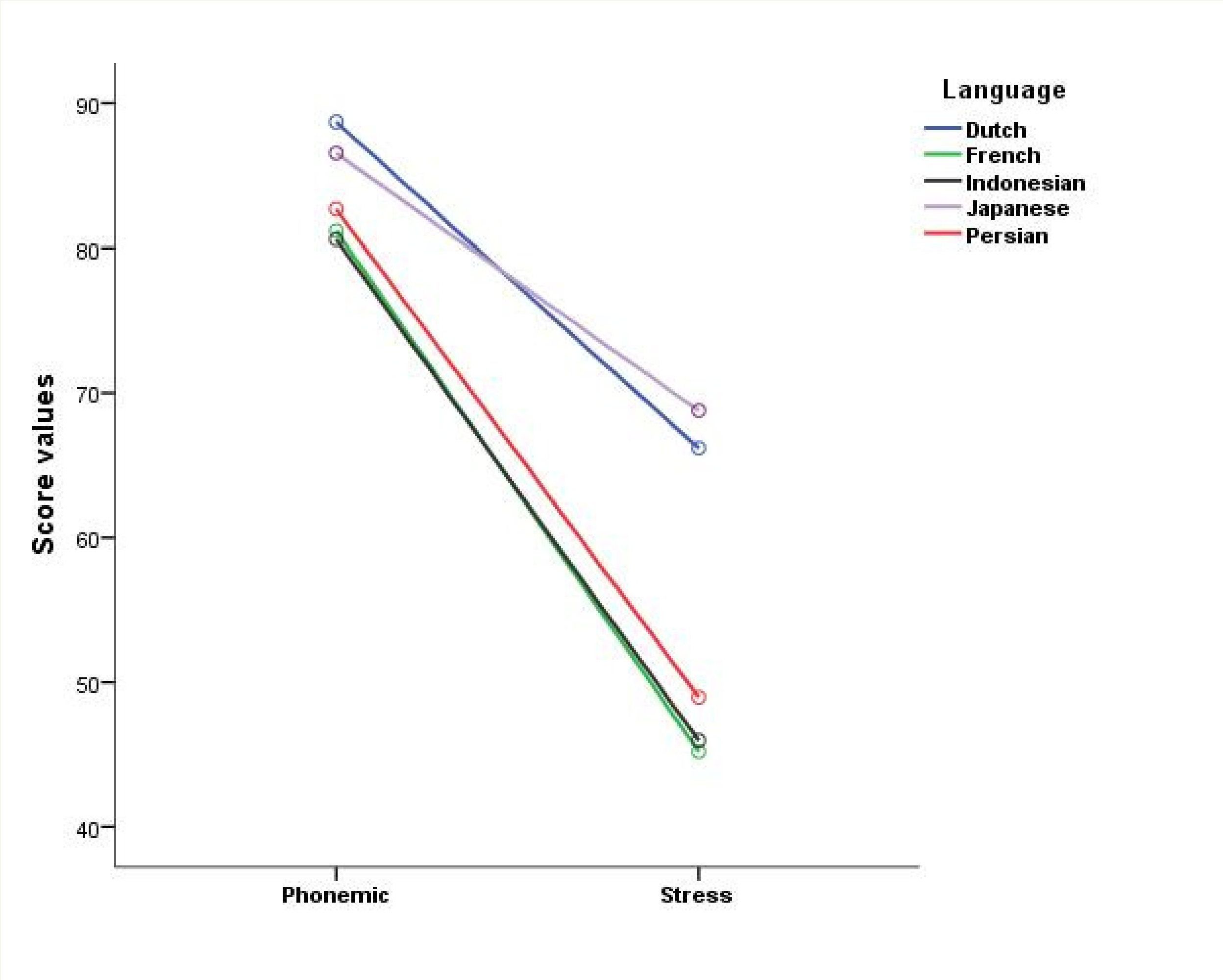
  númi-numí; múnu – múku.
- 3-, 4-, and 5-sequence strings.

Testing Persian word accent in stress deafness paradigm

Rahmani, Rietveld & Gussenhoven (in prep.)

- What's the position of *Persian*?

5 languages, 20 participants per language.


Language
Dutch
French
Indonesian
Japanese
Persian
Score values
90
80
70
60
50
40
Phonemic
Stress

Testing Persian word accent in stress deafness paradigm

Rahmani, Rietveld & Gussenhoven (in prep.)

- Persian is like French and Indonesian, Dutch is like Japanese: two groups of languages.
- Conclusion: success in this test is based on presence of stress/accent marking in lexical stems.
- Persian phonological words are postlexical: clitics are prepositions, auxiliaries, pronouns.
- Verbal accent morphology is exceptionless.

Conclusions

- ‘Stress’ (a) structural (obligatory, culminative)
  (b) phonetic realization (non-f0)
  (c) lexically marked stress/tone
- Typology is about classifying linguistic features/phenomena (Hyman, *passim*)

Features

	Phonetic stress	Tone	Obligatory	Culminative	Stem marking
Nigerian English	Yes	Yes	Yes		Initial/ peninitial
Ambonese Malay					
Zwara Tamazight	Yes		Yes	Yes	Loan words
Persian		Yes	Yes	Yes	

Thank you!

Abolhasanizadeh, Vahideh, Mahmood Bijankhan, Carlos Gussenhoven (2012). The Persian pitch accent and its retention after the focus. *Lingua* 122, 1380-1394.

Beckman, Mary E. (1986). *Stress and non stress accent*. Foris.

Bowern, Claire & Erich Round (in prep.) . Yidiny stress, length and truncation revisited.

Eka, David (1985) A phonological study of Standard Nigerian English. PhD Thesis Amadu Bello University, Nigeria.

Gussenhoven, Carlos (to appear). On the intonation of tonal varieties of English. In Markku Filppula, Juhani Klemola & Devyani Sharma (eds) *The Oxford Handbook of World Englishes*.

Gussenhoven, Carlos & Inyang Udofot (2010). Word melodies vs. Pitch accents: A perceptual evaluation of terracing contours in British and Nigerian English. *Speech Prosody 2010*.

Gut, Ulrike (2005) Nigerian English Prosody. *English World-Wide* 26, 153-177.

Hyman, 2012. In defense of prosodic typology. *Linguistic Typology* 6: 341–385.

Jones, Daniel (1932 [1918]). *An Outline of English Phonetics*. Third Edition. Leipzig: Teubner.

Maskikit, R. & Gussenhoven, C. (in preparation). No stress, no pitch accent, no prosodic focus: The case of Ambonese Malay.

Odé, Cecilia (1994). On the perception of prominence in Indonesian. In Cecilia Odé, Vincent J. van Heuven & Ellen van Zanten (eds.) *Experimental studies of Indonesian prosody*. Leiden: Rijksuniversiteit te Leiden, Vakgroep Talen en Culturen van Zuidoost-Azië en Oceanië. 27-107.

Prentice, D.J. (1987). Malay (Indonesian and Malaysian). In Comrie, Bernard (ed.) *The world's major languages*. London: Croom Helm. 913-935.

Rahmani, Hamed, Toni Rietveld & Carlos Gussenhoven (in preparation). 'Stress deafness' reveals lexical marking of stress or tone.

Tabain, Marija, Janet Fletcher & Andrew Butcher (2014). Lexical stress in Pitjantjatjara . *Journal of Phonetics* 42: 52-66.

Udofot, Inyang (2007). A tonal analysis of standard Nigerian English. *Journal of the Nigerian English Studies Association* 3, 58–68.

Van Minde , D. (1997). *Malayu Ambong: Phonology, morphology, syntax.* Leiden: CNWS

Xu, Y., Chen, S.-w., Wang, B., 2012. Prosodic focus with and without post-focus compression (PFC): a typological divide within the same language family? *The Linguistic Review* 29, 131-147.

Labeling in the Wild: Crowdsourcing versus Categorical Perception

Mark Hasegawa-Johnson, Jennifer Cole, Preethi Jyothi and Lav Varshney

Modern speech technology and speech science depend on large labeled corpora. These large corpora are collected under three assumptions: speech is perceived in terms of discrete categories; there is consistency within and across labelers in the categories that are perceived in a given speech sample; a language can be reliably labeled only by a first-language speaker (and not by non-natives). Machine learning requires labeled corpora, and labeled corpora require technologically literate labelers, therefore any language without a large paid population of such labelers finds itself ignored by speech technology. Recent machine learning theory provides three mechanisms by which we might alleviate the growing speech technology gap between well-resourced and under-resourced languages. First, active learning is a set of methods in which the machine learning algorithm plays an active role in its own education, e.g., by selecting the speech waveforms whose labels would be most informative. In the best case, theoretical guarantees show that active learning can reduce the error rate of a trained classifier from 1/n to exp(-n), where n is the number of labeled data samples. Algorithms for active learning can be tested on well-resourced languages, and having been proven on known tasks in known languages, can then be deployed in order to rapidly develop technology in under-resourced languages. Second, methods of crowdsourcing allow speech labeling tasks to be distributed to a large number of anonymous labelers, some of whom may be less reliable than others. Finally, we can consider the possibility of mismatched crowdsourcing: the acquisition of speech labels from labelers who are not native speakers of the language under study, or by native speakers who lack adequate linguistic training to perform a desired labeling task. Mismatched labeling is a kind of lossy communication channel, according to which some of the information in the original signal has been systematically deleted by the untrained ears of the labeler. As in the case of any other lossy channel, the problem can be solved by a low bit-rate auxiliary channel: in this case, a small amount of data transcribed by a trained expert native speaker may be enough to recover all of the information lost by the mismatched crowdsourcing experiment. In the paradigms of active learning and matched crowdsourcing, theoretical performance guarantees specify upper bounds on the error rates that result from labeling with any given number of labelers. The paradigm of mismatched crowdsourcing has been much less heavily studied, either in theory or in practice, but work in multilingual speech recognition, rapid prosody transcription, and second-language pronunciation error detection suggests the possible shape of future results.

Labeling in the Wild: Crowdsourcing versus Categorical Perception

Mark Hasegawa-Johnson, Jennifer Cole, Preethi Jyothi and Lav Varshney

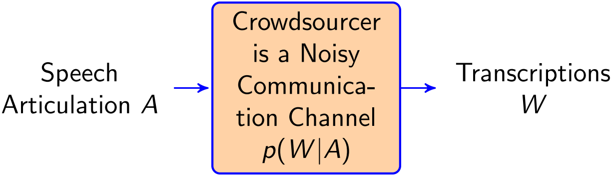
University of Illinois

LabPhon 2014

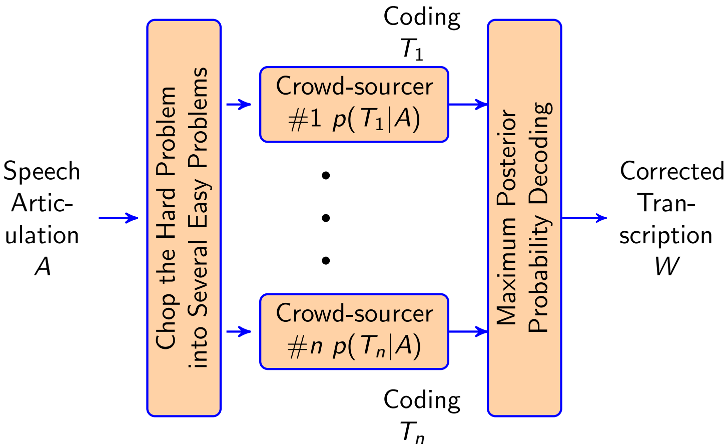
Main Points of This Talk

1. **Crowdsourcing** can give you data cheaply, but crowdsourcers make mistakes. Majority voting reduces error, but triples (or worse) your cost.
2. **Error-correcting codes:** If you factor each hard question into several easy (binary) questions, you can improve accuracy more cheaply, because each crowdsourcer only needs to be partially correct.
3. **The science of easy questions:** Factoring a hard problem into easy problems allows you to find out what linguistically naïve crowdsourcers think about hard linguistic questions.
4. **Crowdsourcing versus categorical perception:** Transcription in the wrong language introduces errors. The errors can be modeled using FST models of transcriber cognition.

The Main Problem: Crowdsourcing Introduces Noise



Proposed Solution: Chop the Hard Problem into Several Easy Problems



Outline

Outline

Sample Problem

Speech recognition fails for Betelgeusians because **they have two heads,** which results in an unusual pronunciation of their vowels. To solve this problem, we would like to learn a classifier that can distinguish /i/ from /e/.

1. Both classes Gaussian w/Identity covariance.
2. Gaussian mixture models (GMM) w/Identity covariance.

A Famous Betelgeusian

(Zaphod Beeblebrox, *Hitchhikers' Guide to the Galaxy*)

Assume Random Training Data, $\mathcal{D}_0$ and $\mathcal{D}_1$

Randomly choose a training sample

$$\begin{aligned} \text{From} \quad \text{Class} \quad & 0: \\ \mathcal{D}_0 \quad = \quad & \{\vec{x}_1, \ldots, \vec{x}_n\} \\ \text{From} \quad \text{Class} \quad & 1: \\ \mathcal{D}_1 \quad = \quad & \{\vec{x}_{n+1}, \ldots, \vec{x}_{2n}\} \end{aligned}$$

Each $\vec{x}$ is a d-dimensional vector, e.g., cepstrum. Estimate the sample means

$$\hat{\mu}_0 = \frac{1}{n}\sum_{i=1}^{n} \vec{x}_i, \quad \hat{\mu}_1 = \frac{1}{n}\sum_{i=n+1}^{2n} \vec{x}_i$$

Assume a Fixed Testing Datum, $\vec{x}$

$g(\vec{x})$ is the classifier function, e.g.,

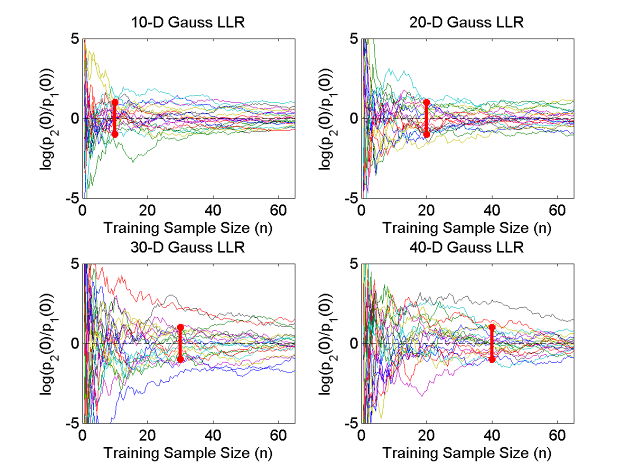
$$g(\vec{x}) = \frac{|\vec{x} - \hat{\mu}_0|^2 - |\vec{x} - \hat{\mu}_1|^2}{2}$$

$$= \vec{x}^T(\hat{\mu}_1 - \hat{\mu}_0) + \frac{|\hat{\mu}_0|^2 - |\hat{\mu}_1|^2}{2}$$

For **random** $\hat{\mu}_0$, $\hat{\mu}_1$ and **fixed** $\vec{x}$, $g(\vec{x})$ is a Gaussian plus the difference of two scaled χ^2 random variables:

$$\sigma_{g(\vec{x})} = \sigma_x \sqrt{\frac{d}{n}} \sqrt{\frac{2|\vec{x}|^2}{d} + \sigma_x^2}$$

d-Dim Classifier Converges Like d Variances



$g(\vec{0})$ as a function of n, multiple random trials. $\sigma_{g(\vec{0})} = \sigma_X^2 = 1$ when $n = d$.

Unit Variance Gaussian Mixture Model (GMM)

Test Rule

$$g(\vec{x}) = \ln\left(\frac{\sum_{k=1}^{m}\mathcal{N}\left(\hat{\mu}_{1k}, I\right)}{\sum_{k=1}^{m}\mathcal{N}\left(\hat{\mu}_{0k}, I\right)}\right)$$

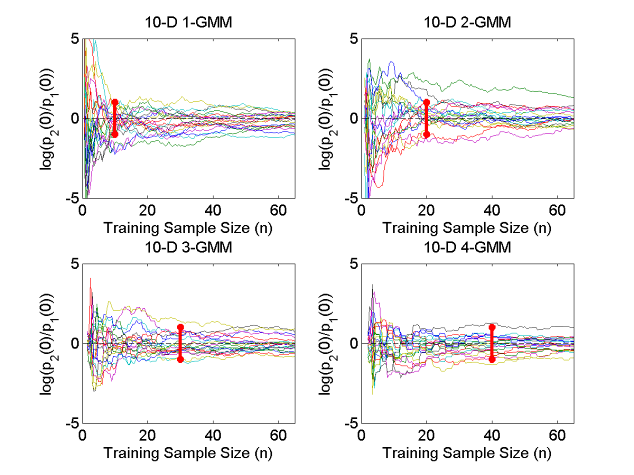
Training Rule

$$\hat{\mu}_{ck} = \frac{1}{n_{ck}}\sum_{\vec{x}_i\in\mathcal{D}_{ck}}\vec{x}_i, \quad 0\le c\le 1,\ 1\le k\le m,\ \sum_{k=1}^{m} n_{ck} = n$$

If $\vec{x}$ is fixed but $\mathcal{D}_{ck}$ are random, then $g(\vec{x})$ is random

$$\frac{g(\vec{x})}{2m\sigma_x^2/n} \sim \chi^2(d), \qquad \sigma_{g(\vec{x})} \approx \sigma_x^2\sqrt{\frac{md}{n}}$$

m-GMM Converges like m Gaussians



$g(\vec{0})$ as a function of n, multiple random trials. $\sigma_{g(\vec{x})} = \sigma_x^2 = 1$ when $n = md$.

How Many Labeled Data Are Needed?

To learn a d-dimensional m-GMM, with a classifier function $g(\vec{x})$ that has standard error at most $\epsilon\sigma^2$, we need

$$n \geq \frac{md}{\epsilon^2}$$

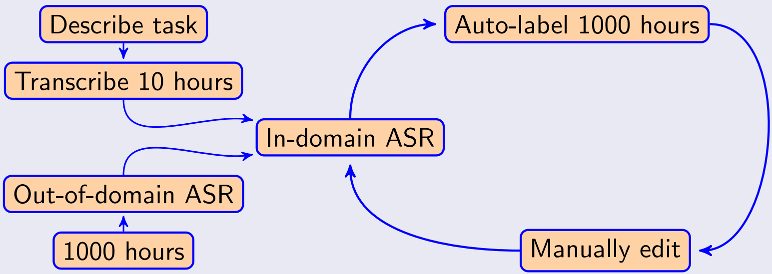
For example, to train a 6-GMM for 40-dimensional cepstra so that $\sigma_{g(\vec{0})} \leq 0.1\sigma_X^2$ requires

$$n \geq 24,000$$

example cepstra (4 minutes of speech) per phone.

- If we have 40 phones represented by exactly 4 minutes of speech per phone, that's 160 minutes (2.67 hours of speech).
- If we have 5000 context-dependent triphones, we need 200,000 minutes (3500 hours).

The Speech Technology Development Cycle



Assumptions

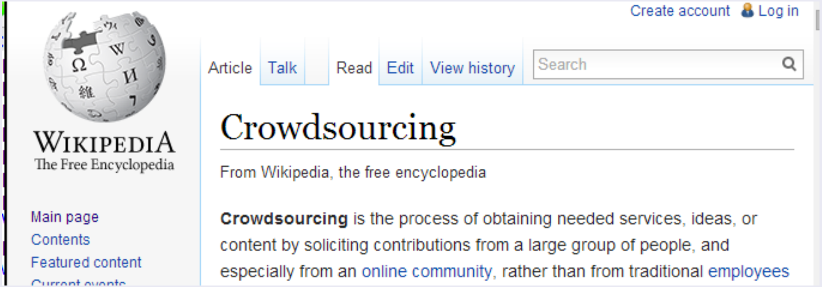
1. Speech is perceived in terms of discrete phonological categories
2. Labelers perceive those categories consistently, as long as. . .
3. Labelers must be drawn from a homogenous linguistic community.

Who are the Labelers?

Source	Motivation	Speed @ Wage
Academic	High	$20 \frac{\text{transcriber hours}}{\text{speech hour}}$ @ \$25/hour
Professional	High	$6 \frac{\text{trainscriber hours}}{\text{speech hour}}$ @ \$30/hour
Crowd	Variable	$600 \frac{\text{hits}}{\text{speech hour}}$ @ \$0.1/hit

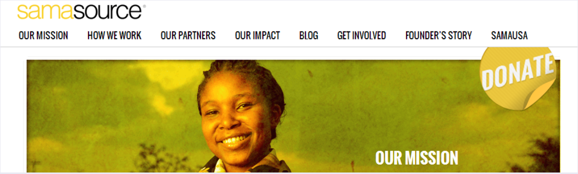
- Cieri et al., "The Fisher Corpus: a Resource for the Next Generations of Speech-to-Text," LREC 2004
- Eskenazi et al., *Crowdsourcing for Speech Processing*, 2013

Crowdsourcing

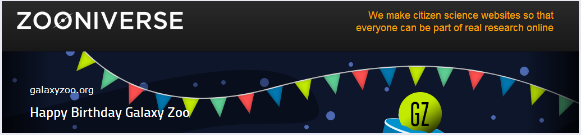


Crowdsourcing sites include big companies. . .

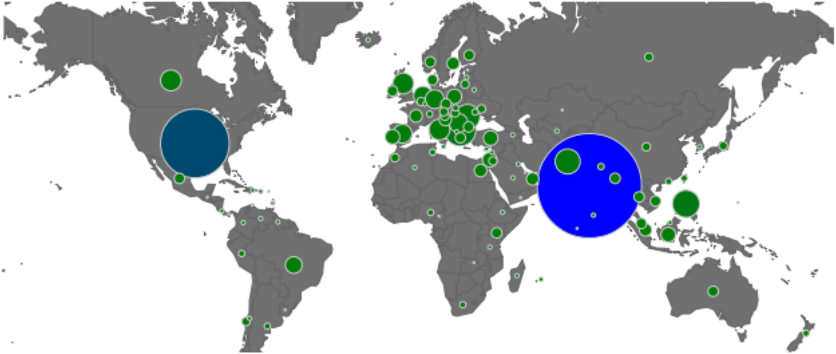
. . . international development organizations. . .



. . . and scientific consortia.



The Language Demographics of Mechanical Turk (Pavlick, Post, Irvine, Kachaev and Callison-Burch, 2013)



Number of workers per country, based on geolocating the IP addresses of 4983 workers. India: 1998, US: 866, Philippines: 142, Egypt: 25, Russia: 10, Sri Lanka: 4. (Pavlick et al., 2013).

Cost, Speed and Quality (Mason and Watts, 2009)

- Payment affects **quantity** of work performed (and **speed**)
- Unexpectedly, payment doesn't affect **quality** of work performed.

Who Turks? (Pavlick et al., 2013)

- USA: mostly people who want a part-time job with scheduling flexibility
- India: mostly full-timers, treat it as a consulting job

Quality Control Methods (Parent, 2011)

1. Before Data Acquisition
   - **Manual,** e.g., choose only workers with good reputation.
   - **Automatic,** e.g., ask a gold standard question, and allow to continue only those who pass.
2. During Data Acquisition (e.g., majority voting)
3. After Data Acquisition
   - **Manual,** e.g., ask other crowdsourcers to validate questionable input.
   - **Automatic,** e.g., get many responses to same question, compare similarity using string edit distance, eliminate outliers

Outline

Main Points of This Talk

1. **Crowdsourcing** can give you data cheaply, but crowdsourcers make mistakes. Majority voting reduces error, but triples (or worse) your cost.
2. **Error-correcting codes:** If you factor each hard question into several easy (binary) questions, you can improve accuracy more cheaply, because each crowdsourcer only needs to be partially correct.
3. **The science of easy questions:** Factoring a hard problem into easy problems allows you to find out what linguistically naïve crowdsourcers think about hard linguistic questions.
4. **Crowdsourcing versus categorical perception:** Transcription in the wrong language introduces errors. The errors can be modeled using FST models of transcriber cognition.

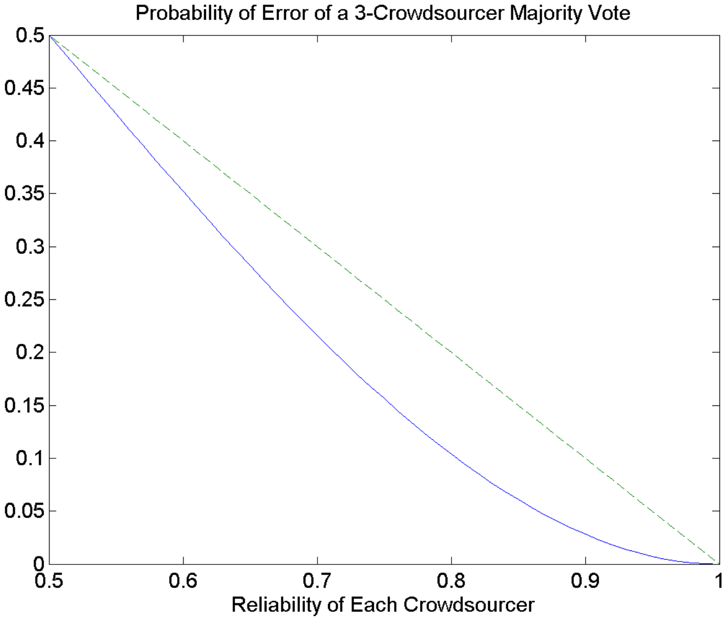
Majority Voting

- Majority voting: assign the same task to ℓ different crowdsourcers. Label the datum with the majority opinion.
- System fails if the majority is wrong. If each crowdsourcer is correct with probability p, then the probability of error is

$$\mathbb{P}_{\text{Error}} = \sum_{k=1}^{\ell/2} \left(\frac{\ell!}{k!(\ell-k)!} \right) p^k (1-p)^{\ell-k}$$

- For example, with $\ell = 3$,

$$\mathbb{P}_{\text{Error}} = 3p(1-p)^2 + (1-p)^3$$


Probability of Error of a 3-Crowdsourcer Majority Vote
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0.5
0.6
0.7
0.8
0.9
1
Reliability of Each Crowdsourcer

Weighted Majority Voting (e.g., Karger, Oh & Shah 2011)

- a_{ij} = answer that i^{th} crowdsourcer gave in response to j^{th} question. (Binary: $a_{ij} \in \{-1, 1\}$)
- $p_{ij} = \Pr\{\ i^{\text{th}}$ crowdsourcer is correct about the j^{th} question$\}$. $(0 \leq p_{ij} \leq 1)$
- r_{ij} = "reference opinion" used to determine whether or not a_{ij} is correct. $(-1 \leq r_{ij} \leq 1)$

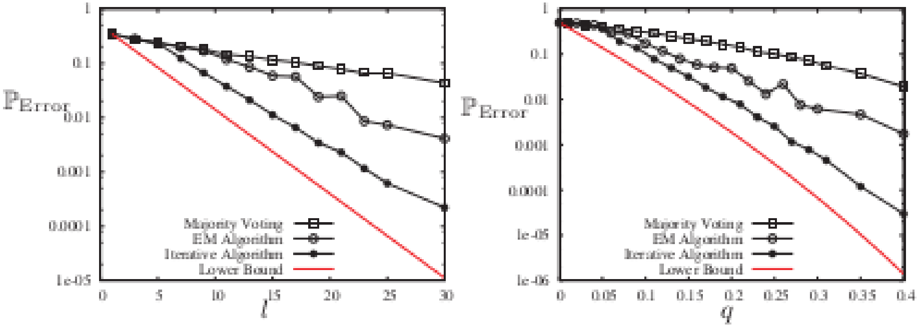
$$r_{ij} \longleftarrow \sum_{k \neq i} a_{kj} p_{kj}$$

$$p_{ij} \longleftarrow \sum_{\ell \neq j} a_{i\ell} r_{i\ell}$$

Iterate until convergence, then compute $r_j = \text{sign}\left(\sum_i a_{ij} \hat{p}_{ij}\right)$, the answer to the j^{th} question.

Weighted Majority Voting is better than Majority Voting (Karger, Oh & Shah, 2011)

- **Theoretical result:** $\mathbb{P}_{\text{Error}} \leq e^{-\ell q/\rho^2}$ for
  - $\ell = \#$ crowdsourcers per question
  - $q = E[2p_{ij} - 1]$ =average crowdsourcer reliability
  - $\rho \approx 3$ is a constant term.
- **Empirical result:**



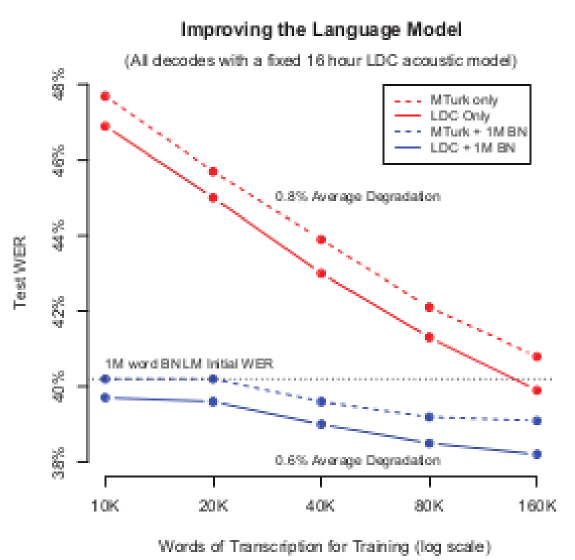
Is Majority Voting Worth the Cost?

Novotney & Callison-Burch (2010) found that

- Training a speech recognizer using crowdsourced transcriptions degrades word error rate (WER) by 2.5%.
- 3-crowdsourcer majority voting results in transcriptions as accurate as LDC, however...
- It's better to have 3$\times$ as much data.
- Benefit of extra data outweighs the cost of increased error.

Is Majority Voting Worth the Cost? Example

WER with varying amounts of language model training data, fixed acoustic model (Novotney & Callison-Burch, 2010, Fig. 2).



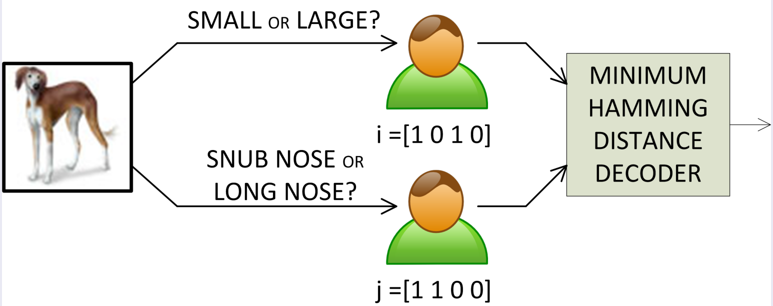
Outline

Main Points of This Talk

1. **Crowdsourcing** can give you data cheaply, but crowdsourcers make mistakes. Majority voting reduces error, but triples (or worse) your cost.
2. **Error-correcting codes:** If you factor each hard question into several easy (binary) questions, you can improve accuracy more cheaply, because each crowdsourcer only needs to be partially correct.
3. **The science of easy questions:** Factoring a hard problem into easy problems allows you to find out what linguistically naïve crowdsourcers think about hard linguistic questions.
4. **Crowdsourcing versus categorical perception:** Transcription in the wrong language introduces errors. The errors can be modeled using FST models of transcriber cognition.

Error Control: Hard Question → Easy Questions (Vempaty, Varshney and Varshney, 2014)

Consider the task of classifying a dog image into one of $M = 4$ breeds: H_0 =Pekingese, H_1 =Mastiff, H_2 =Maltese, or H_3 =Saluki. Crowdsourcers may not be canine experts, but can answer simpler questions.



Easy Questions as a form of Error-Correcting Code

- The "hard question" has M possible answers: $1 \leq m \leq M$. Each is equally likely *a priori*: probability $= \frac{1}{M}$
- "Easy questions" are asked of up to ℓ different crowdsourcers, and they give their answers: $a_j =$ answer given by j^{th} crowdsourcer to whatever question he was asked ($a_j \in \{1, -1\}$)
- $c_{mj} =$ answer he should have given if hypothesis m were correct ("code bit" $c_{mj} \in \{1, -1\}$)

Decoding Rule: Choose $\hat{m}$ for

$$\hat{m} = \arg \min_{1 \leq m \leq M} \sum_{j=1}^{\ell} |a_j - c_{mj}|$$

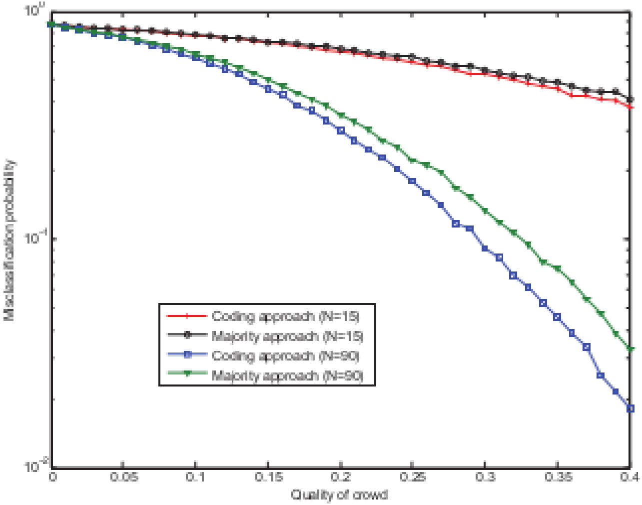
Error-Correcting Code Beats Majority Voting because Even a Wrong Crowdsourcer is Right About Some Things

- Each crowdsourcer answers *easy questions* as though he believes m is the answer to the *hard question*.
- Let $p = \Pr\{\text{crowdsourcer is right about the hard question}\}$
- Let $\frac{1-p}{M-1} = \Pr\{\text{crowdsourcer chooses any particular wrong answer } i \neq m,\ 1 \leq i \leq M\}$

$$1 - \mathbb{P}_{Error} =$$

$$\sum_{m=1}^{M} \frac{1}{M} \sum_{\vec{a}:\hat{m}(\vec{a})=m} \left(\prod_{j=1}^{\ell} \frac{1}{2} \left(1 + a_j \left(p c_{mj} + \frac{1-p}{M-1} \sum_{k \neq m} c_{kj} \right) \right) \right)$$

Error-Correcting Code Beats Majority Voting because Even a Wrong Crowdsourcer is Right About Some Things

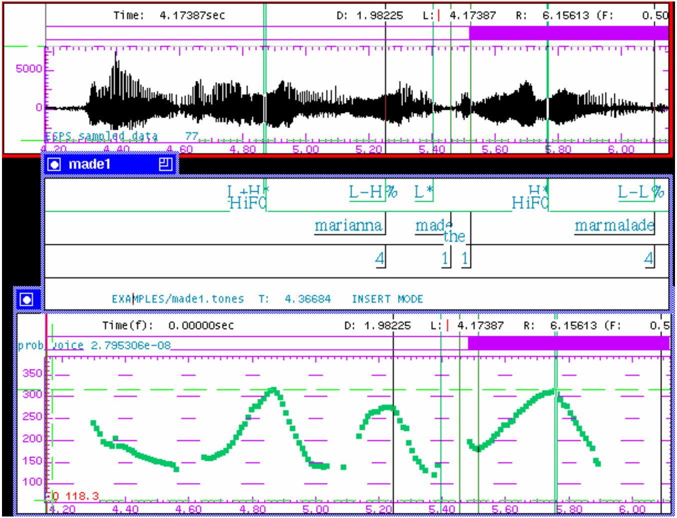


Outline

Main Points of This Talk

1. **Crowdsourcing** can give you data cheaply, but crowdsourcers make mistakes. Majority voting reduces error, but triples (or worse) your cost.
2. **Error-correcting codes:** If you factor each hard question into several easy (binary) questions, you can improve accuracy more cheaply, because each crowdsourcer only needs to be partially correct.
3. **The science of easy questions:** Factoring a hard problem into easy problems allows you to find out what linguistically naïve crowdsourcers think about hard linguistic questions.
4. **Crowdsourcing versus categorical perception:** Transcription in the wrong language introduces errors. The errors can be modeled using FST models of transcriber cognition.

A Hard Problem: Prosodic Phonology (ToBI Guidelines: Beckman & Ayers, 1994)



A Hard Problem: Prosodic Phonology

Different ways in which ToBI has been simplified, in order to simplify the training of automatic prosody detection algorithms. From (Escudero-Mancebo, Gónzalez-Ferreras, Vivaracho-Pascual and Cardeñoso-Payo, 2013)

Classification								
Mapping	H*	H*	H*	H*	high	high	high
	L+H*	L+H*	L+H*	L+H*	high	high	high
	!H*	!H*	H*	!H*	downstepped	downstepped	downstepped
	H+!H*	H+!H*	H+!H*	ignored	high	high	high
	L+!H*	L+!H*	L+H*	ignored	downstepped	downstepped	downstepped
	L*	L*	L*	L*	low	low	low
	L*+H	L*+H	L*+H	ignored	low	low	low
	no label	none	ignored	ignored	unaccented	unaccented	unaccented
	#Classes	8	5	4	4	4	4
	Reference	[7]	[8]	[9]	[10]	[11]	[12]
	Level	word	word	word	syllable	syllable	syllable
	#Words/Syllables	27,767	29,578	28,300	14,599	14,599	14,377
	#Speakers	6	6	6	1	1	1
	Accuracy	**70.8%**	**63.99%**	**56.4%**	**80.17%**	**81.3%**	**87.17%**

[7] González-Ferreras et al. (2012) ; [8] Rosenberg (2010) ; [9] Ananthakrishnan and Narayanan (2008b) ; [10] Ross and Ostendorf (1996) ; [11] Levow (2005) ; [12] Sun (2002)

An Easy Problem: Rapid Prosody Transcription (RPT: Cole, Mo & Hasegawa-Johnson, 2010)

Naïve transcribers: Over 100 UIUC undergraduates, non-experts, performed auditory prosody transcription.

Coarse-grain transcription: Transcribers were given only simple definitions of prominence and boundary, and were instructed to mark words where they heard prominence or boundary.

Strength in numbers: Groups of 15-22 subjects transcribe prosody for the same speech excerpts.

Speed: Transcription is done in real-time, with two listening passes per excerpt, based only on auditory impression.

Rapid Prosody Transcription Example

Vertical bars indicate how the speaker breaks up the text into chunks (boundary)

Underline indicates words that are emphasized or stand out relative to other words (prominence)

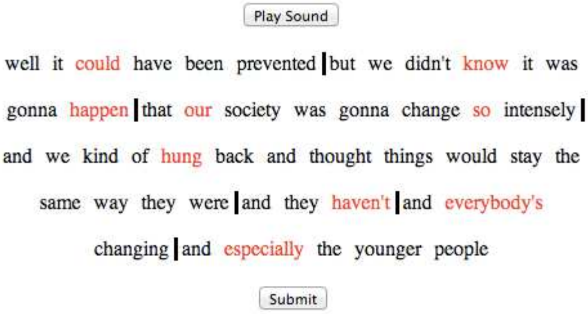
Audio

yeah he's not getting that | I dont think he's getting that | learning | he's | he's more his | that's his grandmother | yknow | watching him. . .

yeah <u>he's</u> not getting that <u>I</u> dont <u>think</u> he's getting that <u>learning</u> he's he's more his <u>that's</u> his <u>grandmother</u> yknow <u>watching</u> <u>him</u>. . .

RPT Example Using LMEDS (LMEDS: Language Markup and Experimental Design Software, Mahrt 2013)

LMEDS screen shot



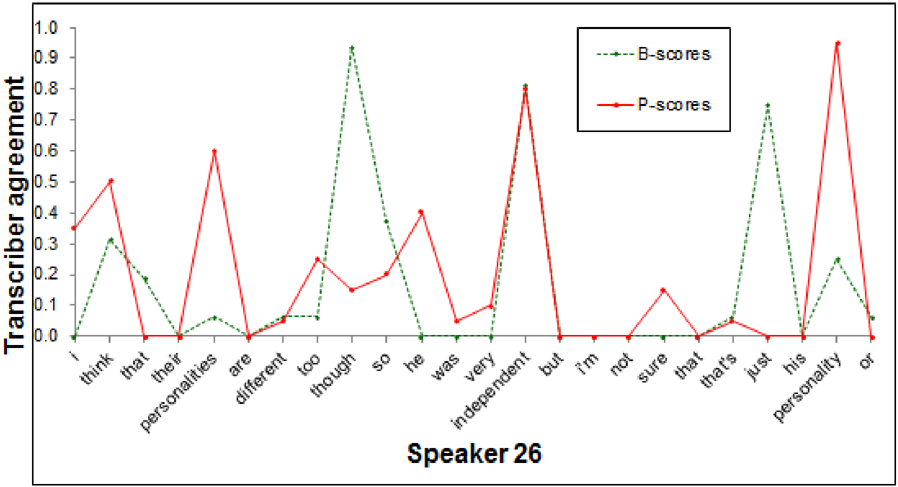
Prosody 'scores'

Each word receives a boundary score (B-score) and a prominence score (P-score).

$$\text{B-score} = T_b/N$$

- T_b = # of transcribers who marked a boundary following that word
- N = total # transcribers
- Similarly, each word receives a prominence score (p-score) indicating how many transcribers marked the word as prominent.

P-scores and B-scores: Example



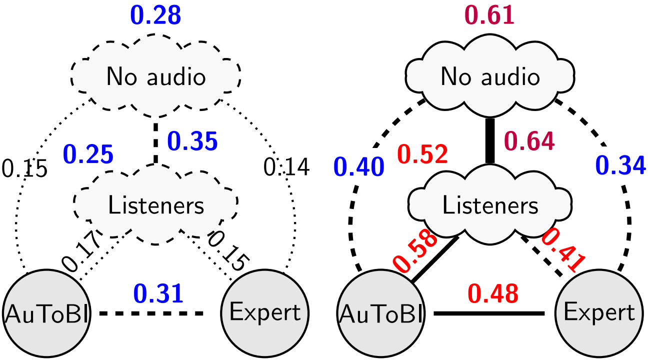
Questions that RPT can ask, but ToBI can't

- Can untrained transcribers label "prosody?"
  Answer: Yes (Cole, Mahrt & Hualde, 2014)
- What are the acoustic and textual correlates of prosodic prominence and boundary, as heard by untrained listeners?
  Some answers: (Cole, Mo & Hasegawa-Johnson, 2010; Cole, Mo & Baek, 2010; Mahrt et al., 2011, 2012)
- Hindi has an F0 movement on each content word, thus English-language models of prominence are largely irrelevant. Does that mean that there is no such thing as prominence in Hindi?
  Results suggest the question is too simple to have a yes/no answer: (Jyothi, Cole & Hasegawa-Johnson, 2014)

An Investigation of Prosody in Hindi Narrative Speech (Jyothi, Cole & Hasegawa-Johnson, 2014)

- Speech data: 10 narrative excerpts in Hindi, about 25 seconds each, from the OGI Multi-language Telephone Speech Corpus
- Transcriptions:
  - RPT with audio: 10 adult speakers of Hindi were asked to mark
    1. how the speaker breaks up the text into chunks (boundary)
    2. words that are emphasized or stand out relative to other words (prominence)
  - RPT without audio
  - ToBI: 1 linguist Ph.D., native speaker of Hindi, ToBI-trained in the USA
  - AuToBI software (Rosenberg, 2010) trained using English-language data

Kappa-score results: Prominence and Boundary



Prominences **Phrase boundaries**

0.1-0.2: Slight agreement

0.2-0.4: Fair agreement

0.4-0.6: Moderate agreement

0.6-0.8: Good agreement

Outline

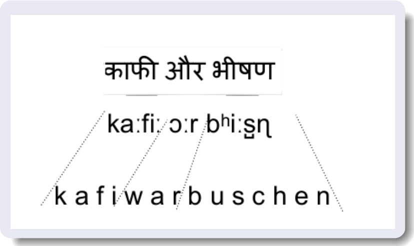
Main Points of This Talk

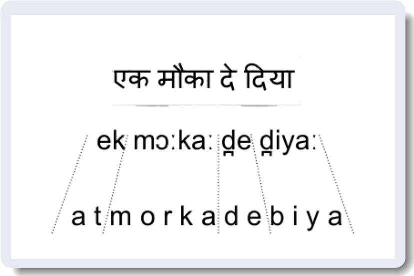
1. **Crowdsourcing** can give you data cheaply, but crowdsourcers make mistakes. Majority voting reduces error, but triples (or worse) your cost.
2. **Error-correcting codes:** If you factor each hard question into several easy (binary) questions, you can improve accuracy more cheaply, because each crowdsourcer only needs to be partially correct.
3. **The science of easy questions:** Factoring a hard problem into easy problems allows you to find out what linguistically naïve crowdsourcers think about hard linguistic questions.
4. **Crowdsourcing versus categorical perception:** Transcription in the wrong language introduces errors. The errors can be modeled using FST models of transcriber cognition.

Mismatched Crowdsourcing: Non-Native Transcription

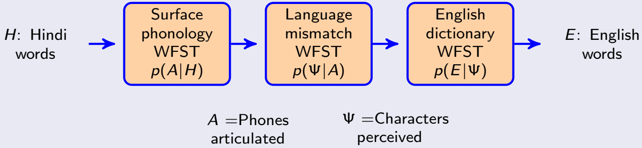
Kalluri vaanil kaayndha nilaavo. . . (Prabhu Deva and Jaya Seal, 2000, as heard by Buffalax=Mike Sutton in 2007)

Examples: Mismatched Transcriptions
Experimental Data, Hindi transcribed as English



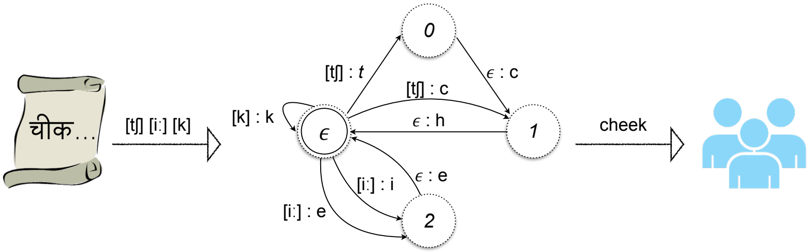


Finite State Transducer Models (I)



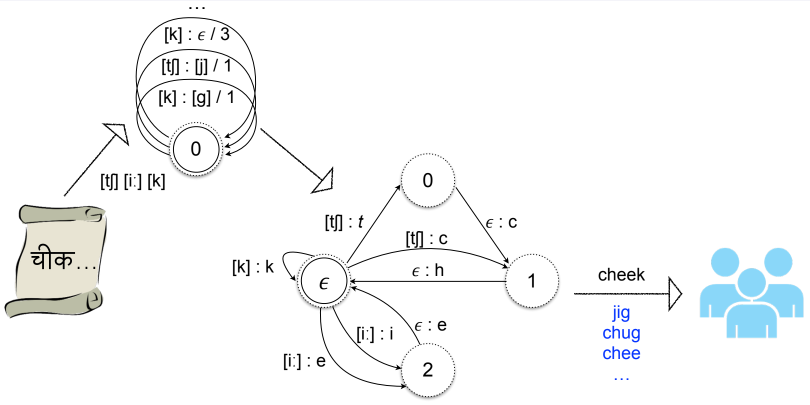
Finite State Transducer Models (II)

Mapping Hindi Words to English Letters



Finite State Transducer Models (III)

Including Hindi Surface Phonology WFST



Estimating the Mismatch FST

In order to estimate the mismatch FST, we need training data.

- A: Fine phonetic transcription by a Hindi-speaking linguist

$$A = [a_1, a_2, \ldots]$$

- Ψ: Ask crowdsourcers to write nonsense syllables instead of English words.

$$\Psi = [\psi_1, \psi_2, \ldots]$$

Mismatched Crowdsourcing: Task Description

Speech materials: Interviews in Hindi from Special Broadcasting Service (SBS, Australia) radio podcasts (mostly spontaneous, formal speech).

Data set: $\approx$ 52 minutes of data excised from speech of 5 interviewers totaling $\approx$ 10K words. Transcribed with phonetic labels by a Hindi expert.

Provided to Mechanical Turk workers: Total of 2074 speech excerpts ($\approx$ 2 secs each) with overlapping 0.5 sec segments. Workers asked to transcribe what they hear using nonsense English syllables.

MTurk worker statistics: Total of 68 workers. 40/68 familiar with English only. Other languages familiar to workers mainly included Spanish, Japanese and Chinese.

Mismatched Crowdsourcing: Task Description

Speech materials: Interviews in Hindi from Special Broadcasting Service (SBS, Australia) radio podcasts (mostly spontaneous, formal speech).

Data set: $\approx$ 52 minutes of data excised from speech of 5 interviewers totaling $\approx$ 10K words. Transcribed with phonetic labels by a Hindi expert.

Provided to Mechanical Turk workers: Total of 2074 speech excerpts ($\approx$ 2 secs each) with overlapping 0.5 sec segments. Workers asked to transcribe what they hear using nonsense English syllables.

MTurk worker statistics: Total of 68 workers. 40/68 familiar with English only. Other languages familiar to workers mainly included Spanish, Japanese and Chinese.

Mismatched Crowdsourcing: Task Description

Speech materials: Interviews in Hindi from Special Broadcasting Service (SBS, Australia) radio podcasts (mostly spontaneous, formal speech).

Data set: $\approx$ 52 minutes of data excised from speech of 5 interviewers totaling $\approx$ 10K words. Transcribed with phonetic labels by a Hindi expert.

Provided to Mechanical Turk workers: Total of 2074 speech excerpts ($\approx$ 2 secs each) with overlapping 0.5 sec segments. Workers asked to transcribe what they hear using nonsense English syllables.

MTurk worker statistics: Total of 68 workers. 40/68 familiar with English only. Other languages familiar to workers mainly included Spanish, Japanese and Chinese.

Mismatched Crowdsourcing: Task Description

Speech materials: Interviews in Hindi from Special Broadcasting Service (SBS, Australia) radio podcasts (mostly spontaneous, formal speech).

Data set: $\approx$ 52 minutes of data excised from speech of 5 interviewers totaling $\approx$ 10K words. Transcribed with phonetic labels by a Hindi expert.

Provided to Mechanical Turk workers: Total of 2074 speech excerpts ($\approx$ 2 secs each) with overlapping 0.5 sec segments. Workers asked to transcribe what they hear using nonsense English syllables.

MTurk worker statistics: Total of 68 workers. 40/68 familiar with English only. Other languages familiar to workers mainly included Spanish, Japanese and Chinese.

Structure of the Mismatch FST

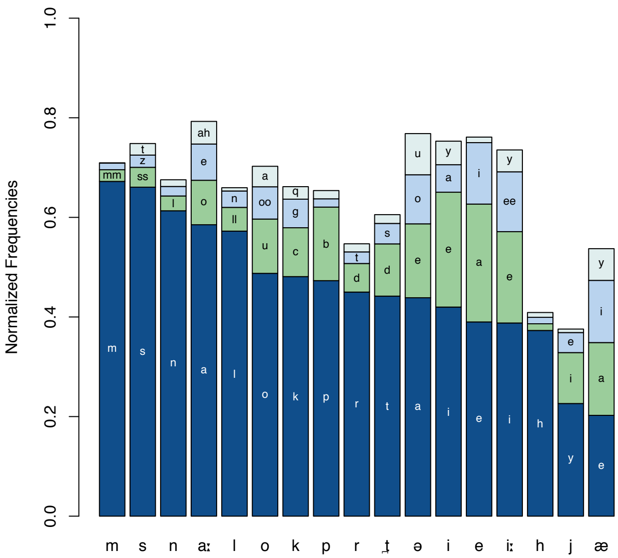
The mismatch FST can be represented as one of these:

- **Distinctive-Feature Weighted Levenshtein Distance:** $-\log p(\Psi|A)$ given by # distinctive feature insertions, deletions, & substitutions from articulated phone string $A = [\ldots, a_t, \ldots]$ to perceived character string $\Psi = [\ldots, \psi_\tau, \ldots]$
- **Learned Levenshtein:** Minimum string-edit distance phone alignment, with substitution costs $\text{SCOST}(a, \psi)$, deletion costs $\text{DCOST}(a)$, and insertion costs $\text{ICOST}(\psi)$ learned from data:

$$-\log p(\Psi|A) \sim \sum_a \sum_\psi \text{SCOST}(a,\psi)\text{NSUBS}(a,\psi)$$

$$+ \sum_a \text{DCOST}(a)\text{NDEL}(a) + \sum_\psi \text{ICOST}(\psi)\text{NINS}(\psi)$$

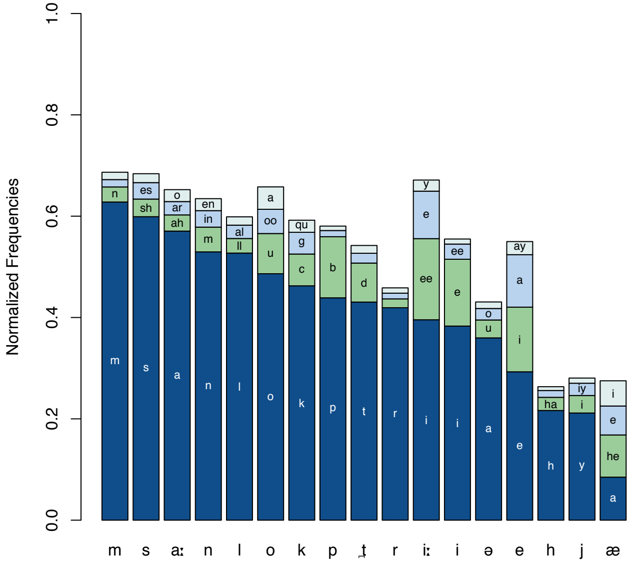
Hindi Sounds Perceived as English Letters (I)



Hindi phones with $\geq$1000 occurrences in the training data

Distinctive Feature-Weighted Levenshtein Mismatch FST: Costs are not learned from data.

Hindi Sounds Perceived as English Letters (II)



Levenshtein Mismatch FST with learned edit costs using the EM algorithm.

Learned Levenshtein Aligns A with Ψ, Allowing us to Compute $p(H|E)$. So what?

Mathematical Theory of Communication (Shannon, 1948)

- Entropy of $H|E$

$$\eta(H|E) = \sum_H \sum_E p(H,E) \log p(H|E)$$

- Perplexity = number of typical inputs given a particular input

$$N(H|E) = 2^{\eta(H|E)}$$

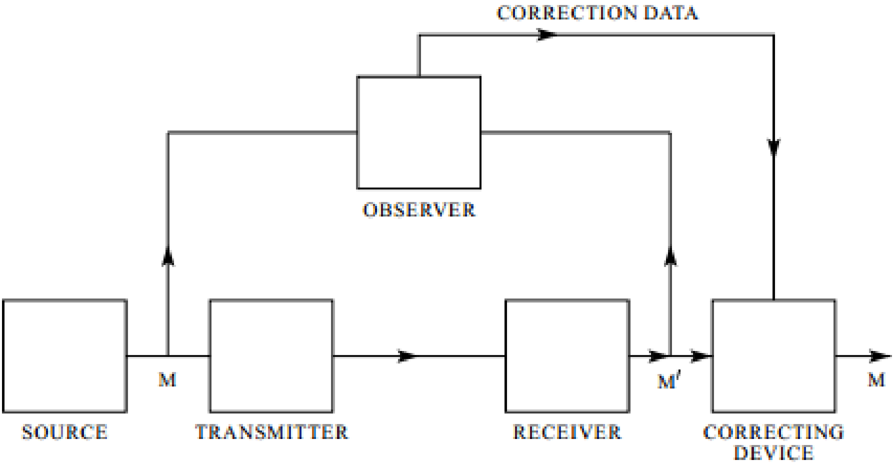
- Shannon, 1948, Theorem 3: As length$(H) \rightarrow \infty$,

$$p(H|E) \rightarrow \begin{cases} \frac{1}{N(H|E)} & H \text{ "typical" given } E \\ 0 & \text{otherwise} \end{cases}$$

Future Work: Post-Editing of Mismatched Crowdsourcing

- Post-editing by a Hindi-speaking linguist
- Prompt screen lists $N(H|E)+1$ options:
  - $N(H|E)$ Hindi sentences that are most probable given the English transcription
  - 1 option that says "OTHER:" allows linguist to type something different
- Scalability via active learning: editor sees only the transcripts with maximum $\eta(H|E)$

Noisy Channel Correction Model (Shannon, 1948, Fig. 8)



Theoretical Result: Bit Rate of the Side Channel

- Hindi language model gives $p(H)$, from which we calculate Entropy $\eta(H|E)$:

$$\eta(H|E) = \sum_{H,E} p(H,E) \ln p(H|E)$$

$$p(H,E) = \sum_{A} \sum_{\Psi} p(E|\Psi) p(\Psi|A) p(A|H) p(H)$$

- Channel capacity of the side channel is

$$C = \log_2 (\# \text{ Correction Options})$$

- Shannon, 1948, Theorem 11 (The "Fundamental Theorem of Communication"):

$$\text{If } C \geq H(H|E) \text{ then } P(\text{ERROR}) \underset{\text{length}(H)\rightarrow\infty}{\longrightarrow} 0$$

Outline

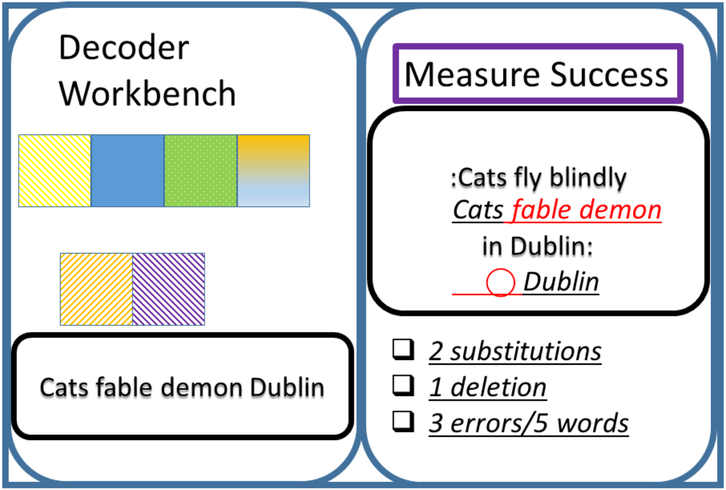
Conclusions

1. **Crowdsourcing** can give you data cheaply, but crowdsourcers make mistakes. Majority voting reduces error, but triples (or worse) your cost.
2. **Error-correcting codes:** If you factor each hard question into several easy (binary) questions, you can improve accuracy more cheaply, because each crowdsourcer only needs to be partially correct.
3. **The science of easy questions:** Factoring a hard problem into easy problems allows you to find out what linguistically naïve crowdsourcers think about hard linguistic questions.
4. **Crowdsourcing versus categorical perception:** Transcription in the wrong language introduces errors. The errors can be modeled using FST models of transcriber cognition.

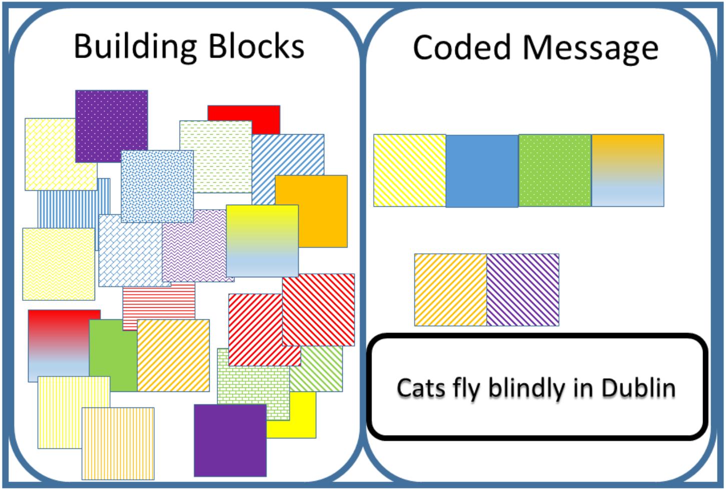
Future Work

- **Further analysis** of the mismatched crowdsourcing model (e.g., "Guessing with side information")
- **Validate** the mismatched crowdsourcing model
- **Scale** using active learning
- **Exploit** mismatched crowdsourcing to build ASR in lots of languages
- **Gamesource** these tasks: write games that bored students will want to play while waiting for the bus.

Example: Secret Agent Game, Decoder Screen



Example: Secret Agent Game, Coder Screen



Thank you!

Infant-directed speech as a window into the dynamic nature of phonology

Reiko Mazuka[1,2], Andrew Martin[1], Yosuke Igarashi[3], Akira Utsugi[4]

1. RIKEN Brain Science Institute
2. Duke University
3. Hiroshima University
4. Nagoya University

Theoretical frameworks of phonology are built largely on the basis of idealized speech, typically recorded in a laboratory under static conditions. Natural speech, in contrast, occurs in a variety of communicative contexts where speakers and hearers dynamically adjust their speech to fit their needs. The present paper demonstrates that phonologically informed analysis of specialized speech registers, such as infant-directed speech, can reveal specific ways segmental and supra-segmental aspects of phonology are modulated dynamically to accommodate the specific communicative needs of speakers and hearers.

Data for the analyses comes from a corpus of Japanese infant-directed speech, consisting of 22 Japanese mothers' spontaneous speech directed to their infant child (Infant-directed speech, IDS) and an adult (adult-directed speech, ADS). The speech samples in the corpus are annotated with segmental, morphological and intonational information. We will show, for example, that the way intonation is exaggerated in Japanese IDS reflects the intonational structure of Japanese, which is different from that of English. We will also demonstrate that rules of phonological grammar, such as devoicing of high vowels and non-high vowels in Japanese, can be differently affected by the needs of the speaker to accommodate the specific characteristics of the listener.

Infant-directed speech as a window into the dynamic nature of phonology

Reiko Mazuka,[1,4] Andrew Martin,[1]
Yosuke Igarashi,[2] & Akira Utsugi[3]

1.RIKEN Brain Science Institute, 2. Hiroshima University,
3. Nagoya University, 4. Duke University

Theme of LabPhone14

Laboratory Phonology beyond the laboratory:
Quantitative analyses of speech produced outside the phonetics laboratory

- Fieldwork-based studies
- Corpus-based approaches
- Acquisition of L1 phonology/prosody

Background

- Real speech occurs in dynamic contexts.
- Speakers adjust their speech dynamically depending on the contexts.
- Phonology needs to account for dynamic aspects of human speech as well.
- Systematic analysis of specialized speech register can offer a window into the dynamic aspects of speech.

Outline

1. RIKEN Japanese mother-infant conversation corpus (R-JMICC)
2. Exaggeration of intonation (Igarashi et al., JASA, 2013)
3. Realization of phonological rule (Martin, et al., Cognition, 2014)

Input for Learning Japanese

- RIKEN-Japanese Mother-Infant Conversation Corpus

  (Mazuka, et al 2006; Igarashi & Mazuka, 2006)
- Participants
  - 22 mothers
    - with their 18-24 month-old infants (12 females, 10 males)
    - From Tokyo area
- Size

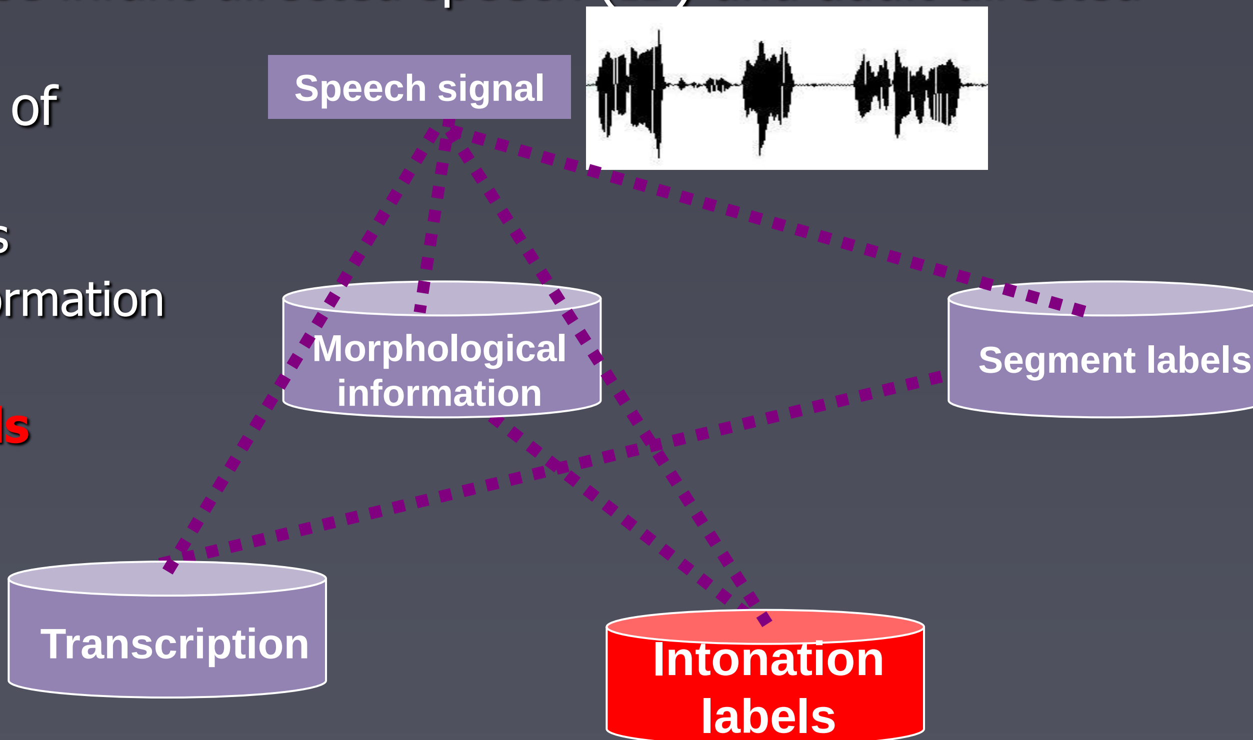
ID	11 hours	50,000 words
AD	3 hours	30,000 words
Overall	14 hours	80,000 words

Tasks (recording environments)

Adult-directed speech	Conversation	Talking with a female experimenter **10 min**	
Infant-directed speech	Book	Playing with picture books **15 min**	
	Toy	Playing with toys **15 min**	

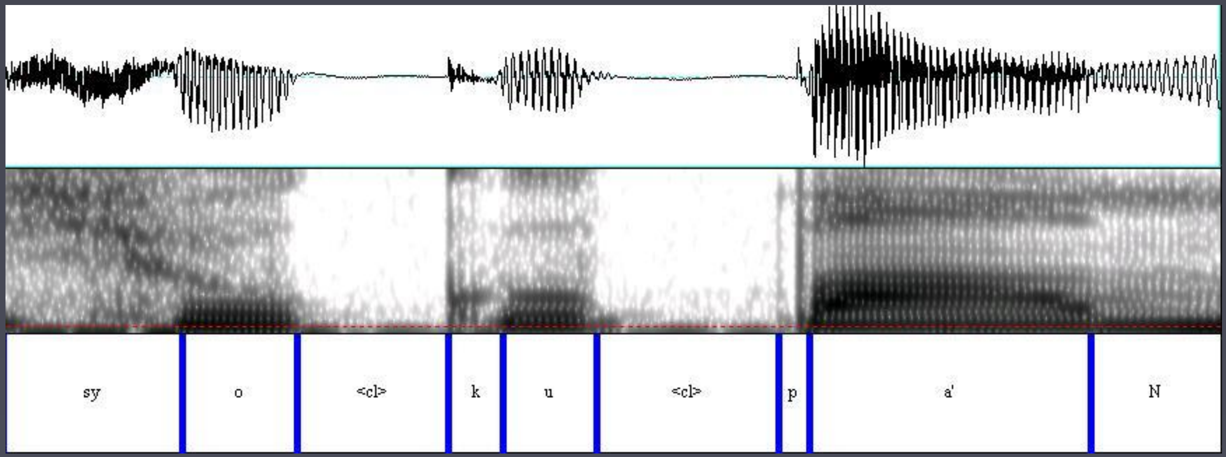
RIKEN Mother-Infant Conversation Corpus

- RIKEN Japanese Mother-Infant Conversation Corpus is a speech database of Japanese infant-directed speech (ID) and adult-directed speech (AD).
- The corpus consists of
  - Speech signals
  - Transcription texts
  - Morphological information
  - Segmental labels
  - **Intonation labels**



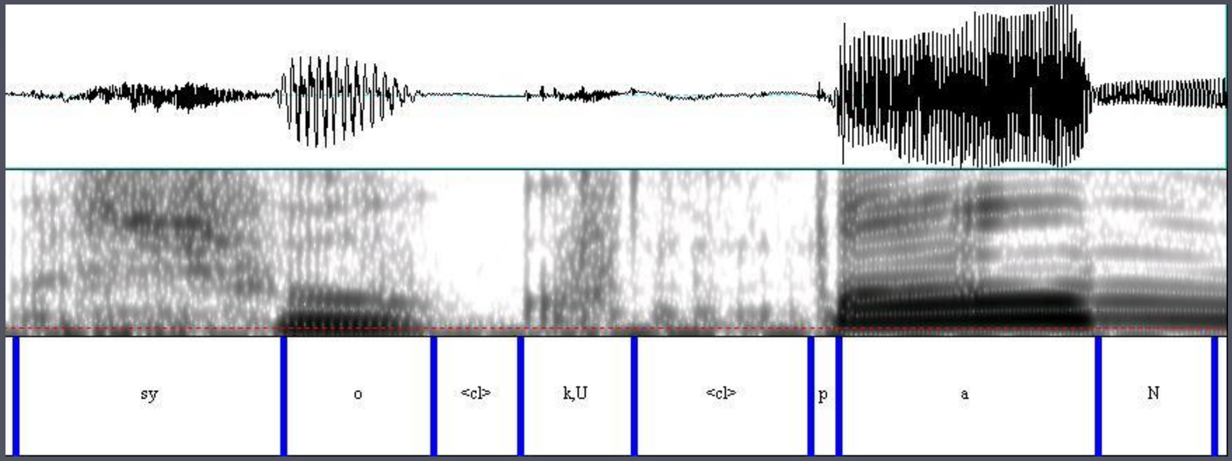
- The intonation labeling is based on the **X-JToBI** scheme (Maekawa et al. 2002, cf. Venditti 2006)
  - It owes its theoretical foundation to the phonological model of Japanese intonation (Pierrehumbert and Beckman 1988).

Segmental labels



‘bread’ without devoicing

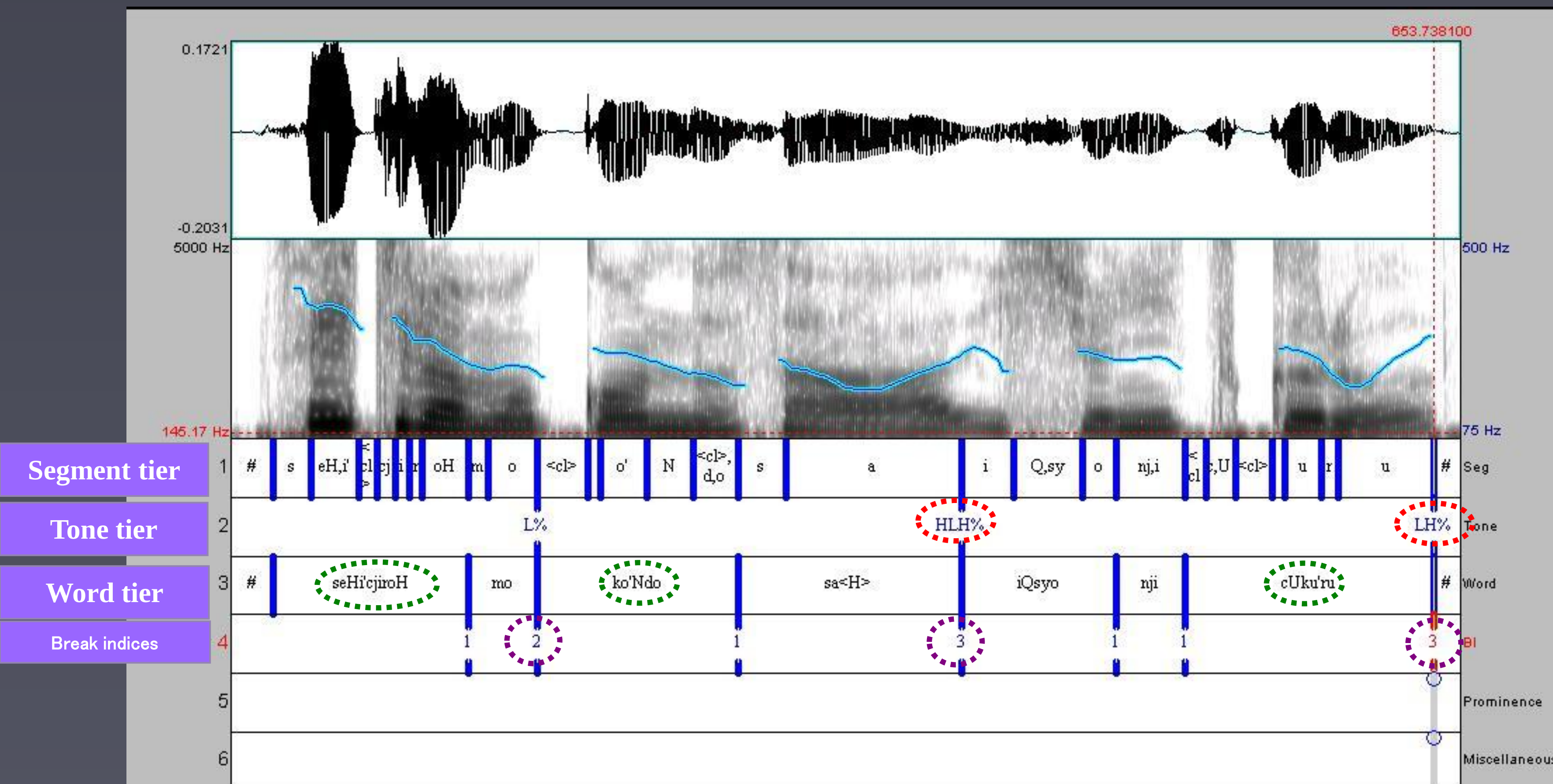
sy o <cl> k **u** <cl> p a N



‘bread’ with devoicing

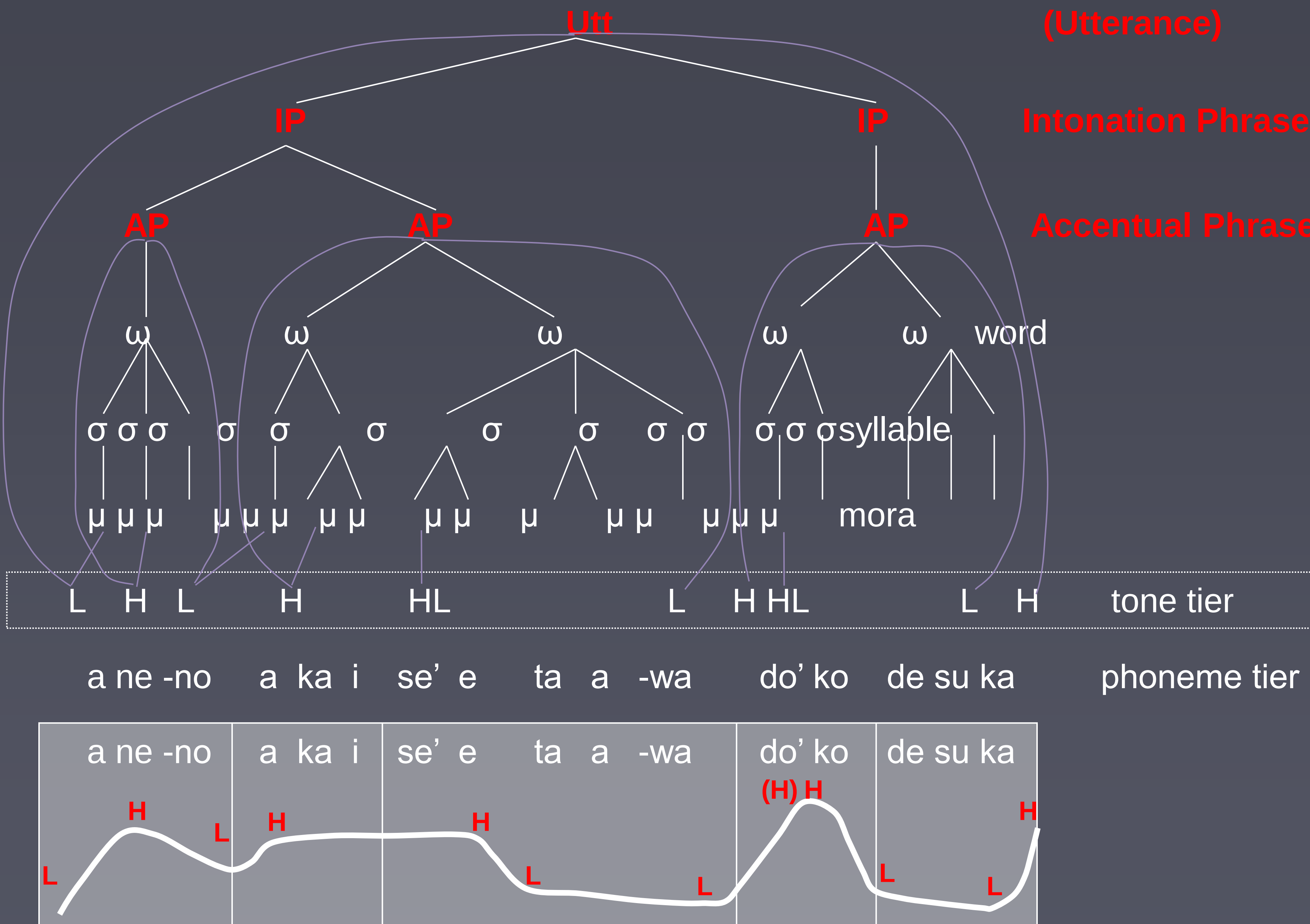
sy o <cl> k **U** <cl> p a N

X-JToBI Coding



「成一郎も今度さあ　一緒に作る？」

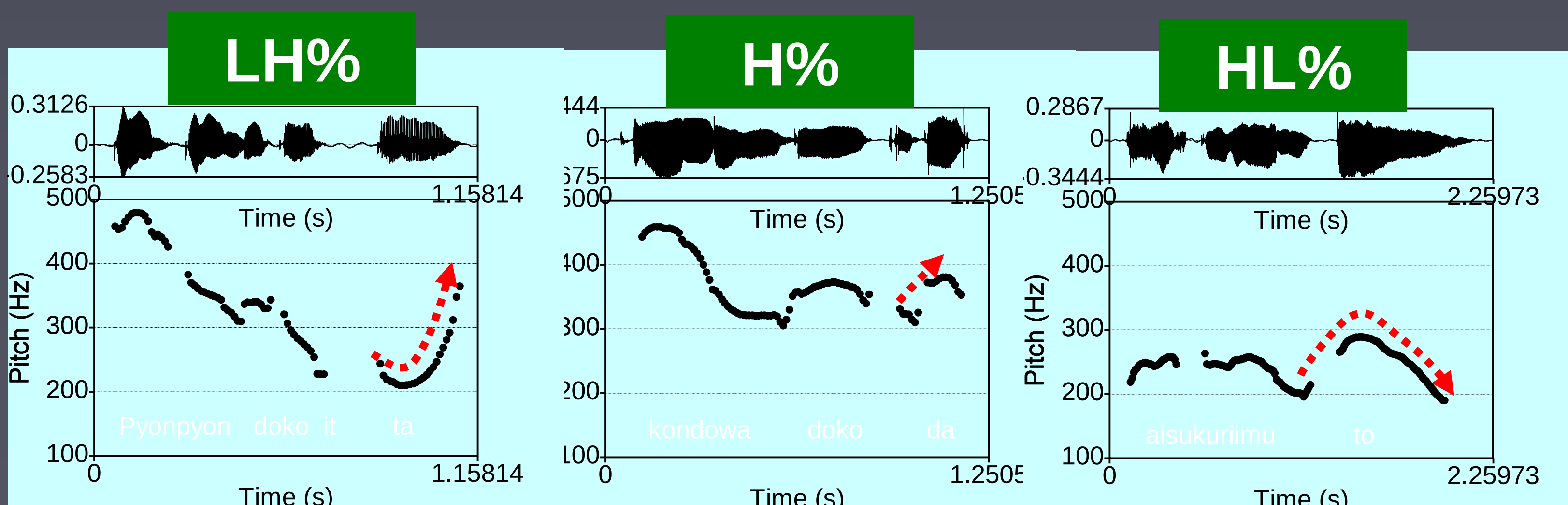
Next time, are you going to make it together, Seiichiro?



'Where is big sister's red sweater?'

What is Boundary Pitch Movement (BPM)?

- BPMs are tonal categories that occur **in the final mora of a prosodic phrase** (typically utterance).
- BPMs contribute to the **pragmatic interpretation** of the utterance, such as questioning and continuation (see Venditti et al. forthcoming).
- Main types of BPMs:



"Where's the frog gone?" "Where is it now?" "An ice cream and…"

Outline

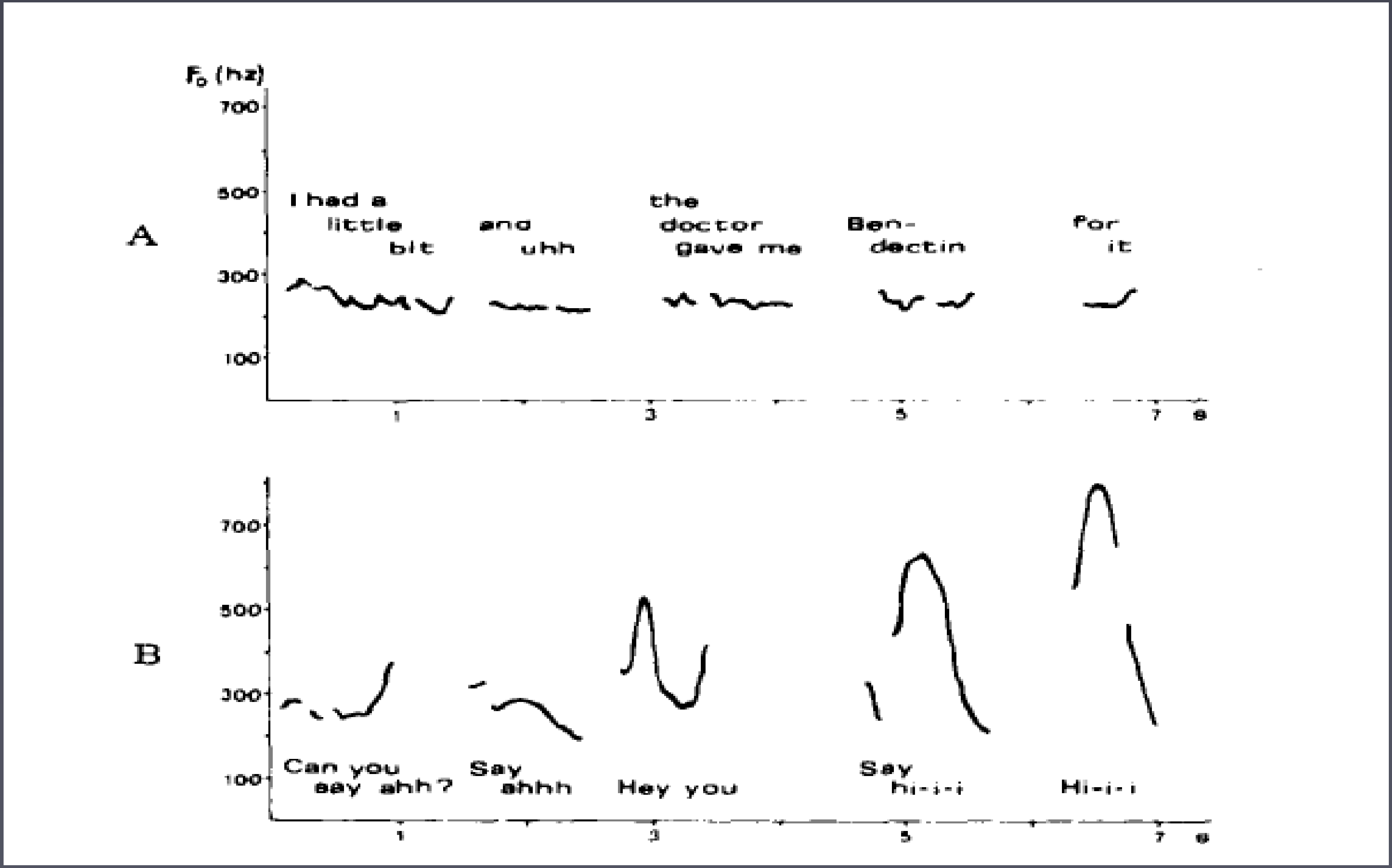
1. RIKEN Japanese mother-infant conversation corpus (R-JMICC)
2. Exaggeration of intonation (Igarashi et al., JASA, 2013)
3. Realization of phonological rule (Martin, et al., Cognition, 2014)

Exaggerated Intonation in IDS?

Igarashi, Nishikawa, Tanaka, & Mazuka, JASA, 2013

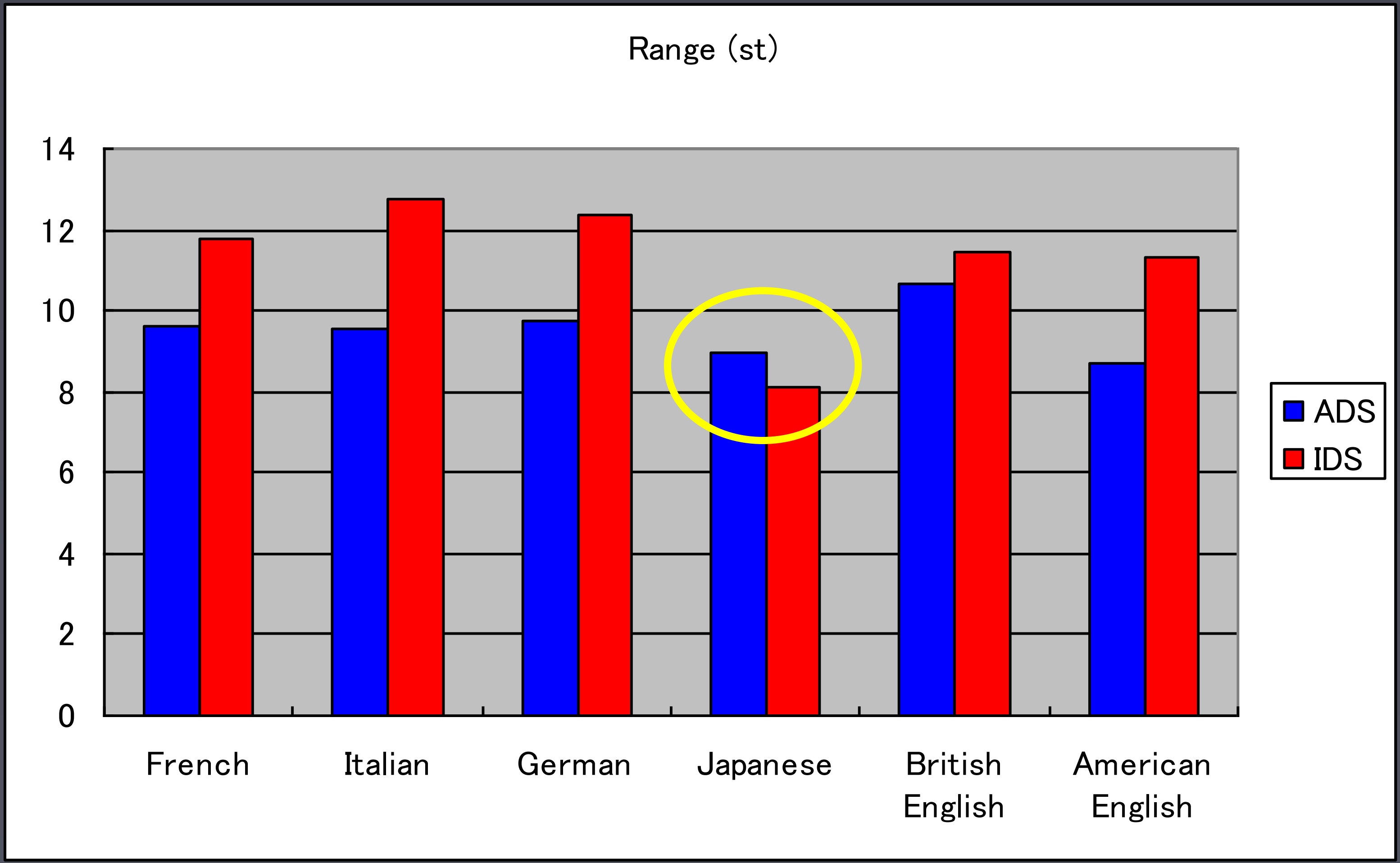
- '*Exaggerated*' *intonation* is one of the most often sited characteristics of IDS prosody (e.g., Fernald et al. 1989).

Pitch Exaggeration in IDS



Fernald (1985)

Fernald et al (1989) found Japanese IDS showed no pitch expansion.



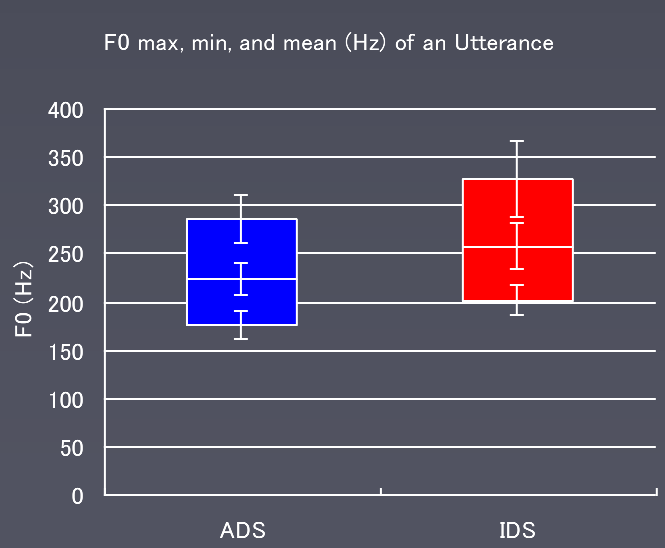
Exaggerated Intonation in IDS?

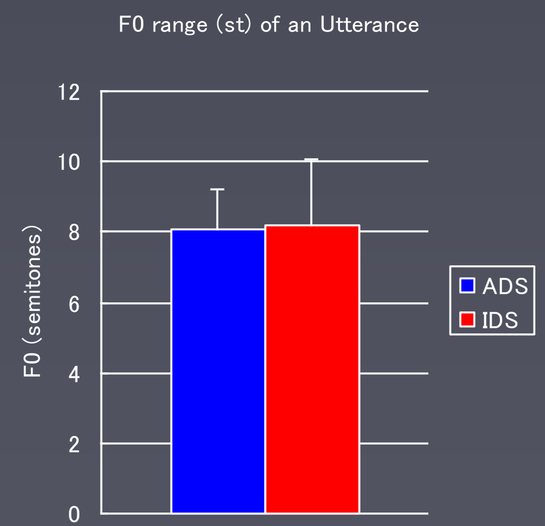
- ‘*Exaggerated*’ *intonation* is one of the most often sited characteristics of IDS (e.g., Fernald et al. 1989).

- Based on *physical measurements* of overall fundamental frequency (F0) contours; e.g., ‘expanded pitch range’, ‘higher pitch level.’
- No reference to the *linguistic structure* of a language’s intonation.
- No exaggeration found in Japanese IDS.

Analysis of R-JMICC Utterance (Overall)

- When we examine the whole utterance
  - Max, min, mean: AD < ID
  - Range (semi tone): AD = ID





Numerial replication of Fernald et al (1989)

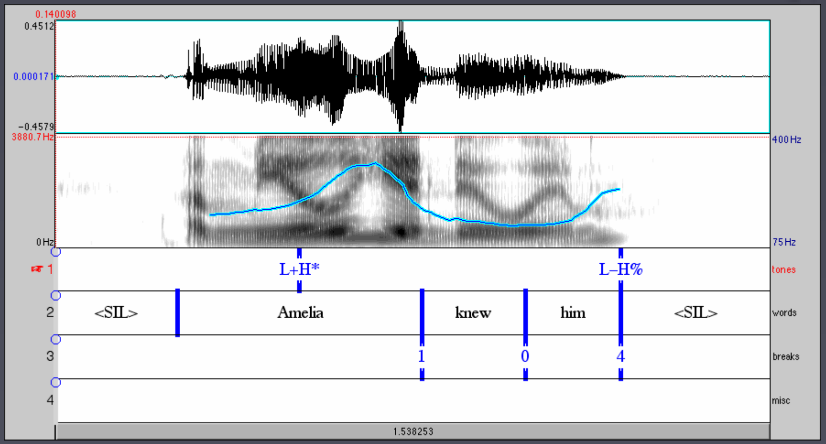
Why?

- Intonation of a language is NOT merely a physical parameter.
- Intonation has a linguistically organized internal structure.
  (e.g. Pierrehumbert 1980; Ladd 1996)
- Examining IDS intonation with reference to the prosodic system of Japanese may reveal language specific nature of Japanese IDS intonation.

Prosodic system of English

A sample taken from:
http://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-911-transcribing-prosodic-structure-of-spoken-utterances-with-tobi-january-iap-2006/index.htm

- A stress, which is lexically specified, involves larger intensity and longer duration.
- Two major components of intonation
  - Pitch accent: appears on a stressed syllable
  - Boundary tone : appears at a phrase edge



Ex.) Amelia knew him.

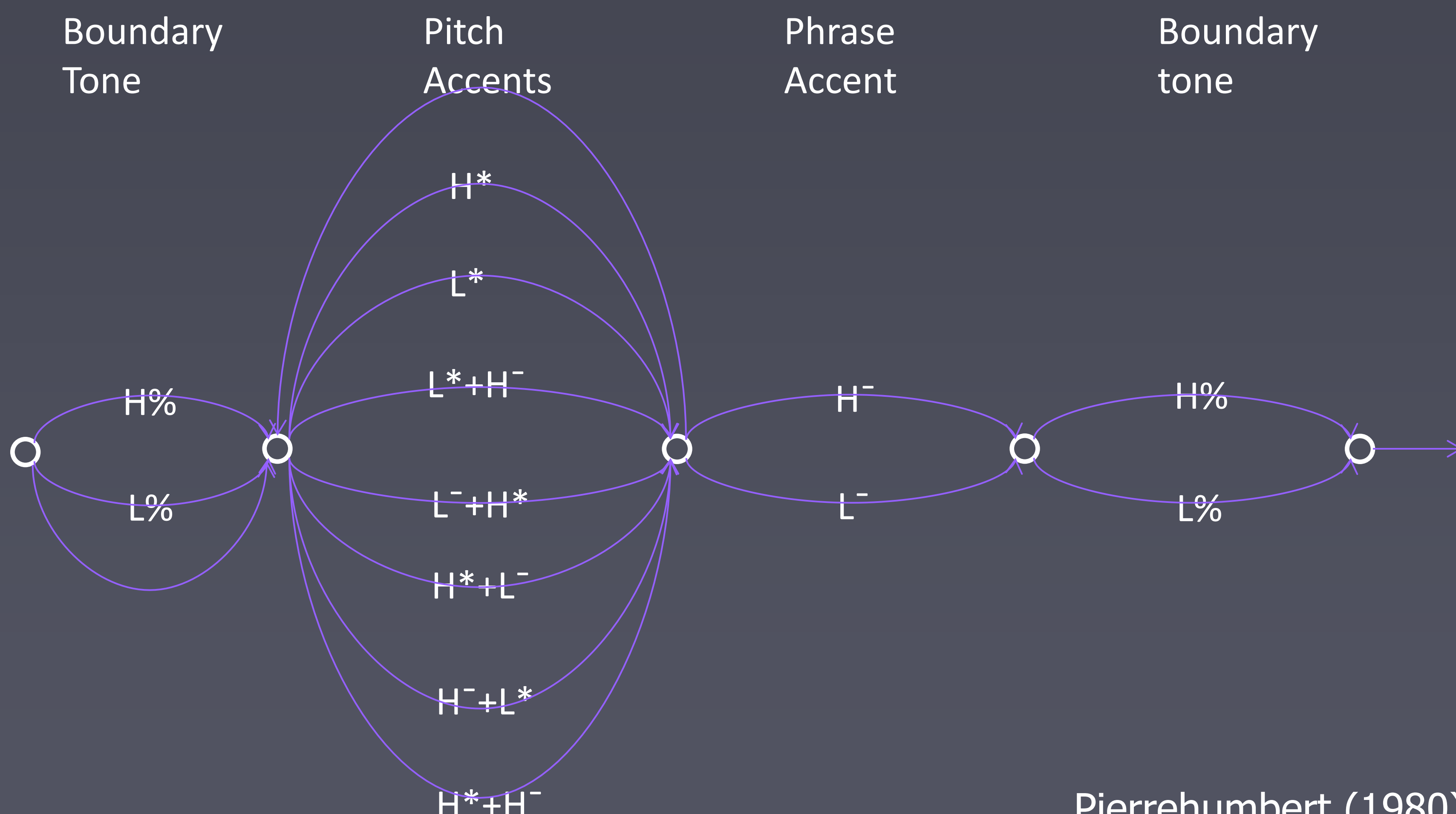
Stress ● ●

Intonation L+H* L-H%

Pitch Accent

Boundary Tone

Finite State Grammar to generate all tunes in English



Pierrehumbert (1980)

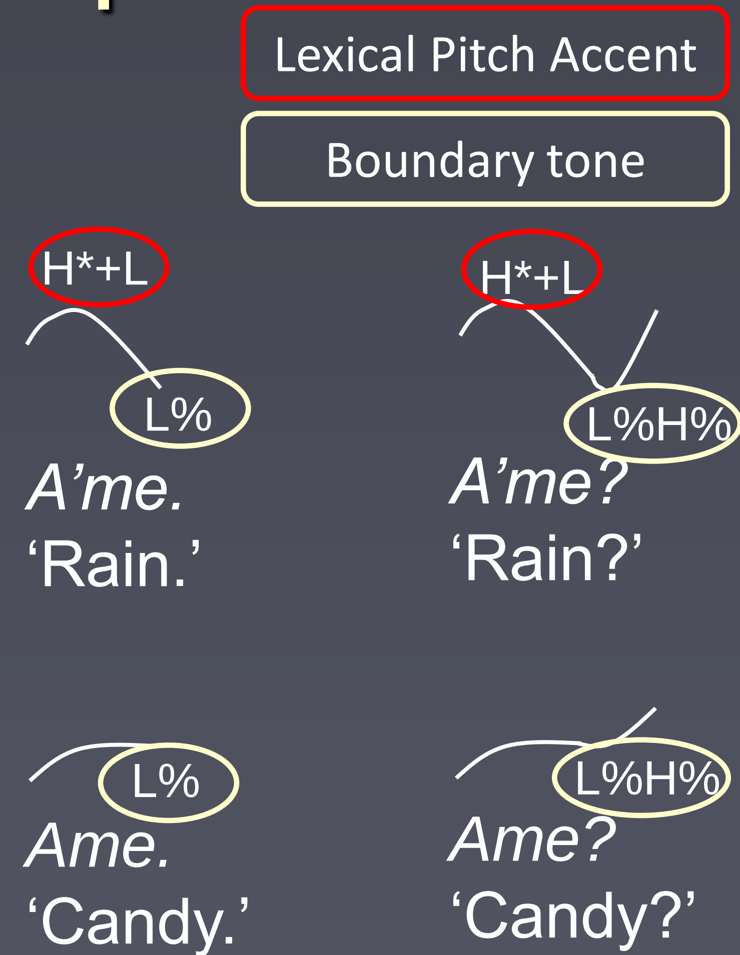
But

- In a language with the edge-prominent prosody, like Japanese & Korean, pitch movement tend to occur phrase finally (e.g., Jun, 2005).

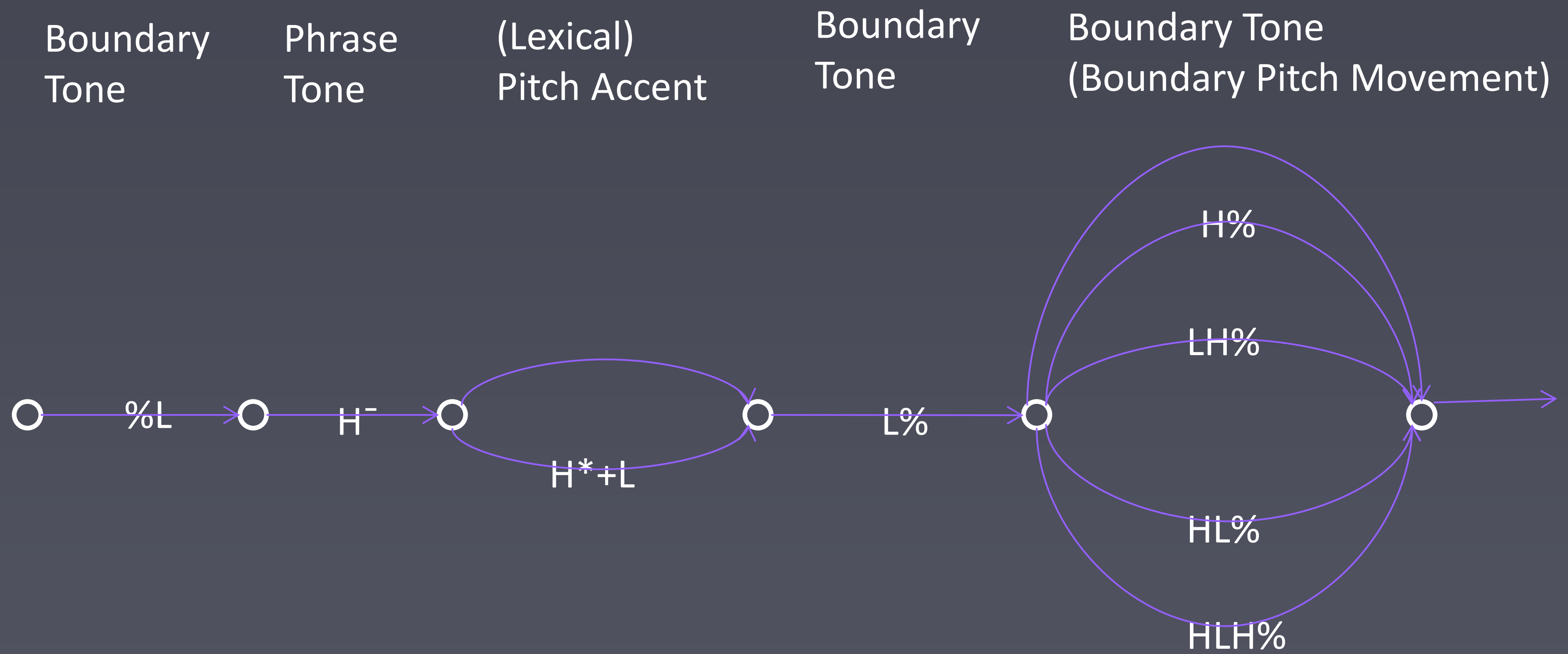
- Boundary pitch movement (BPM)
- LH% -- rising intonation. Question
- HL% -- falling intonation. Turn taking

Prosodic system of Japanese

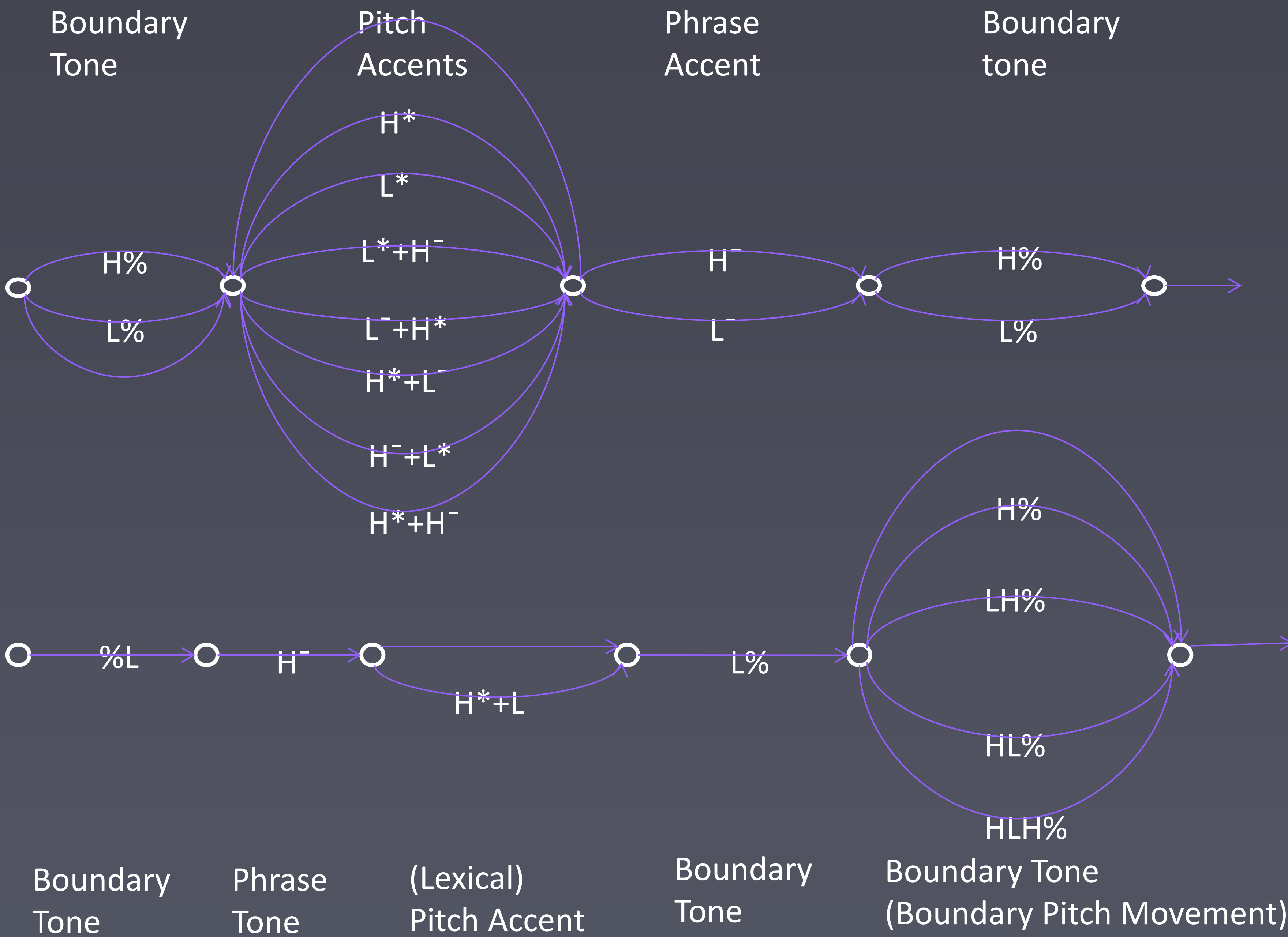
- A word has a lexically specified pitch shape.
  - Accented word: has the H*+L lexical pitch accent
  - Unaccented word: has no lexical pitch accent
- Intonation is realized through boundary tones



Finite State Grammar to generate all tunes in Japanese

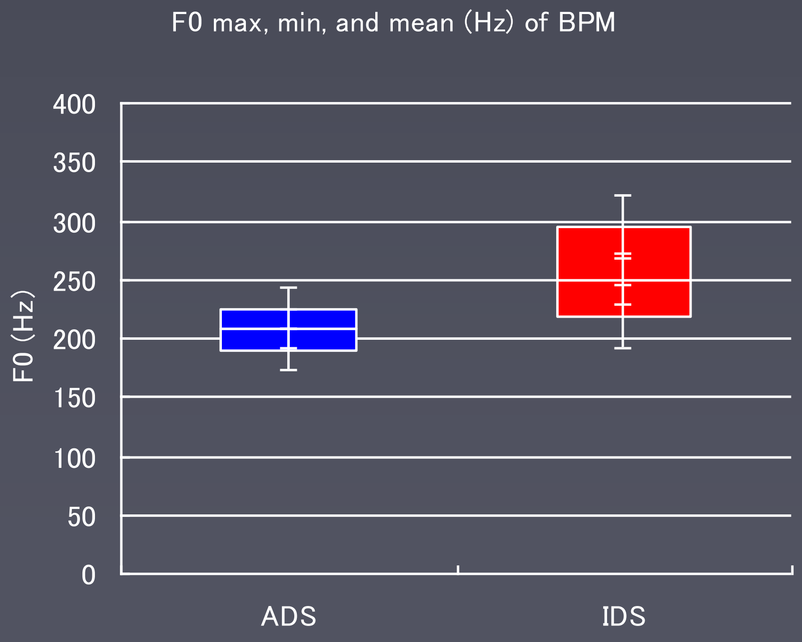


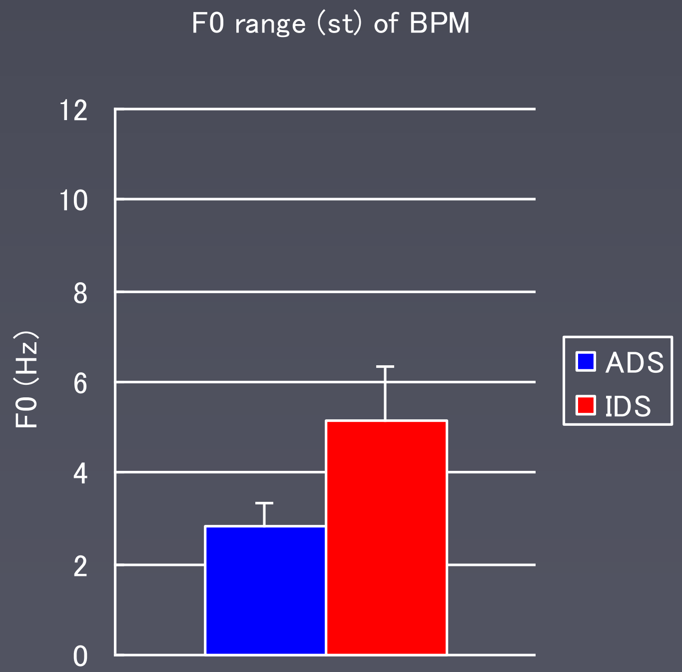
Pierrehumbert & Beckaman (1988), Maekawa et al. (2002)


Boundary Tone
Pitch Accents
Phrase Accent
Boundary tone
H%
L%
H*
L*
L*+H-
L-+H*
H*+L-
H-+L*
H*+H-
H-
L-
H%
L%
%L
H-
H*+L
L%
H%
LH%
HL%
HLH%
Boundary Tone
Phrase Tone
(Lexical) Pitch Accent
Boundary Tone
Boundary Tone (Boundary Pitch Movement)

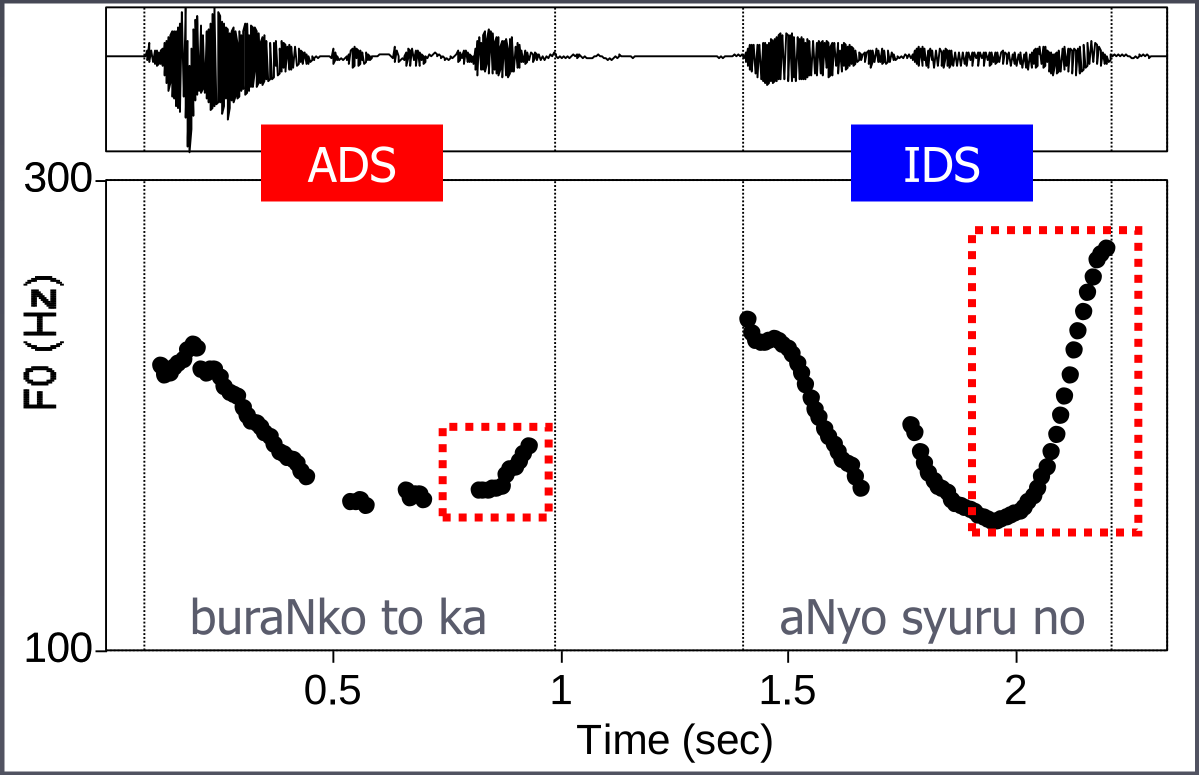
Boundary Pitch Movement

- Syllables with BPM
  - Max, min, mean, range (st): AD < ID

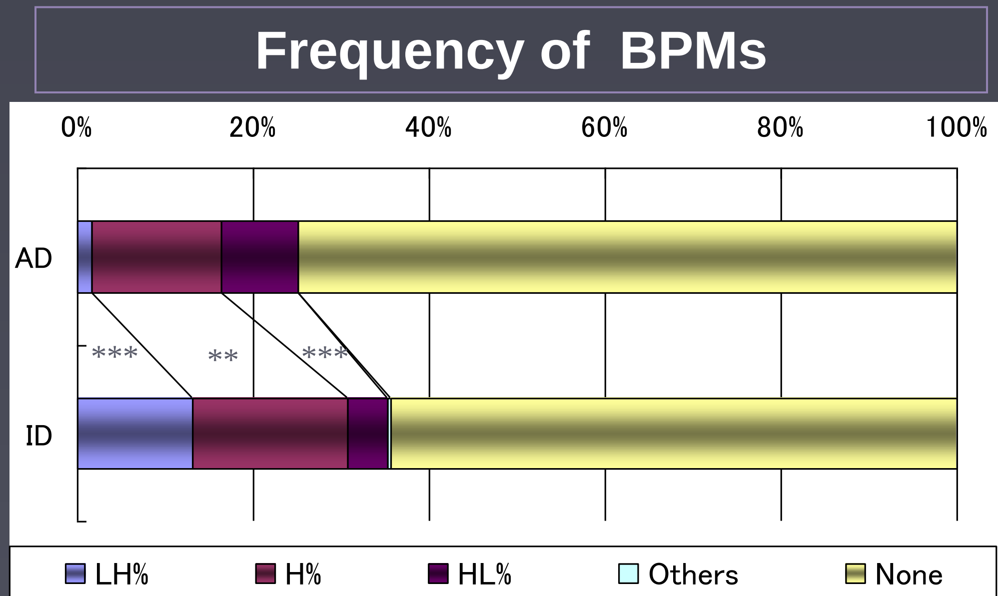




Expansion occurs at BPM

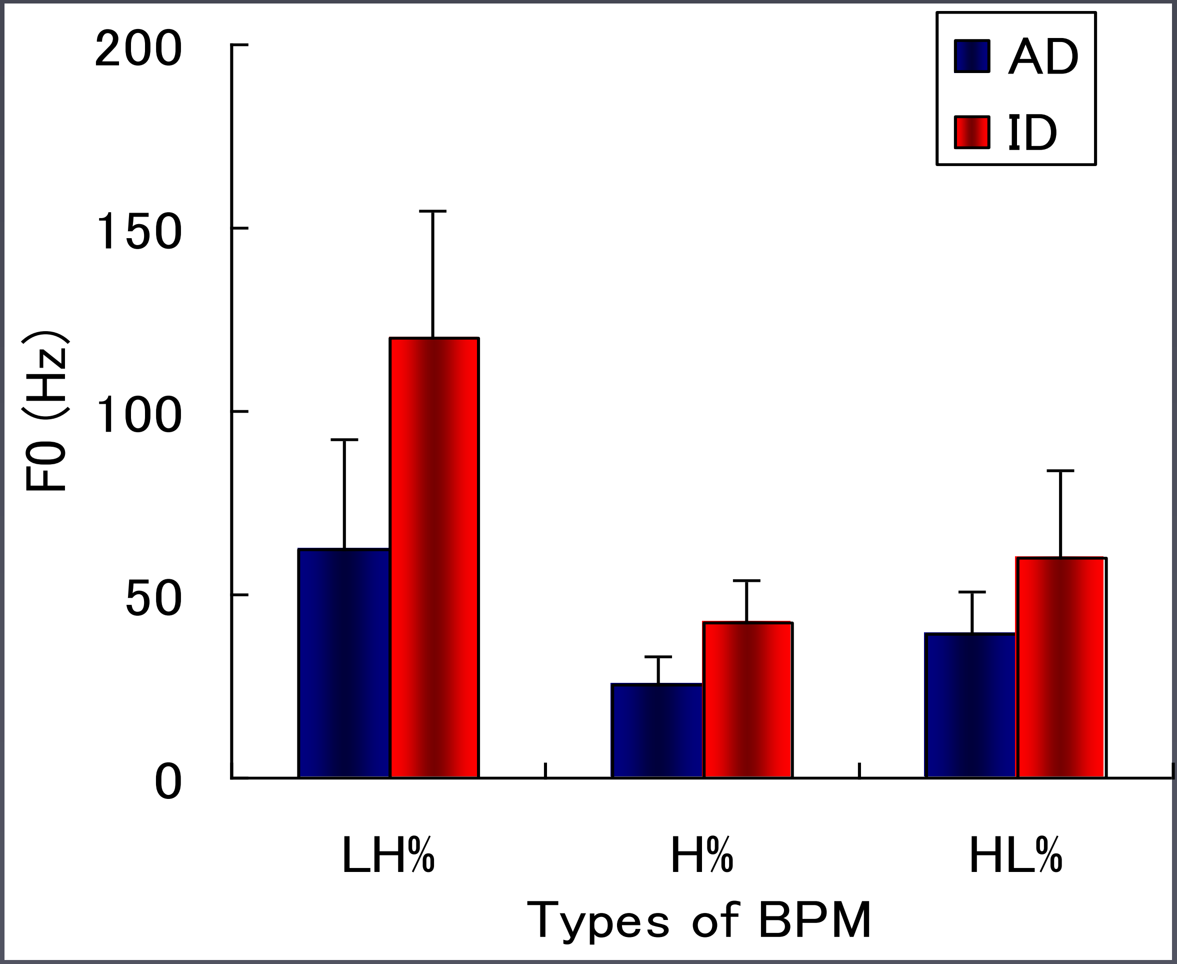


Boundary pitch movements (BPMs)



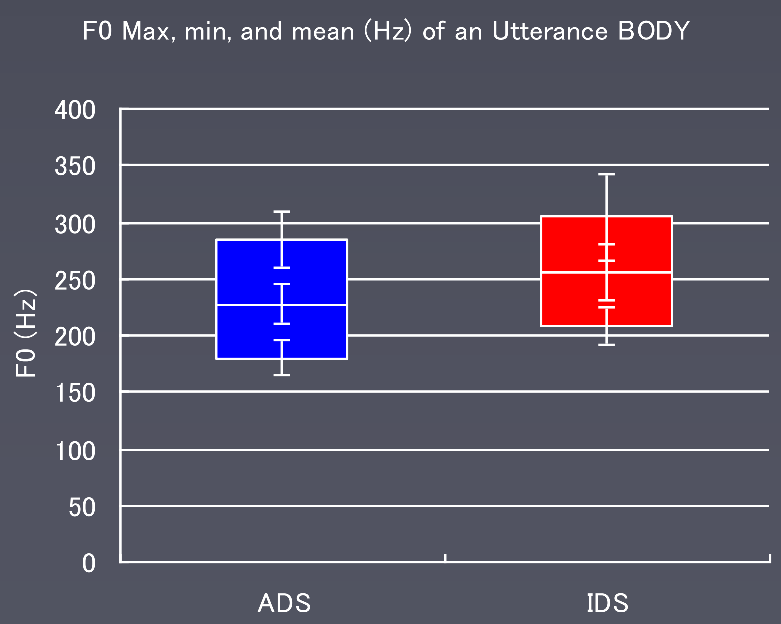
- **More frequent LH% (rise**) in ID.
  - Presumably, more questioning in ID.
- More frequent HL% (rise-fall) in AD.
  - Presumably, more continuation (turn-keeping) in AD

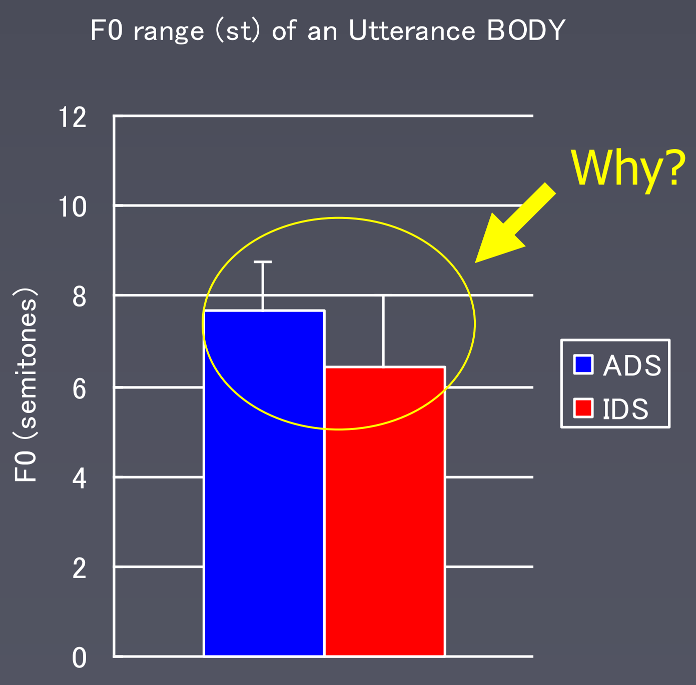
Pitch range is expanded in every BPM type



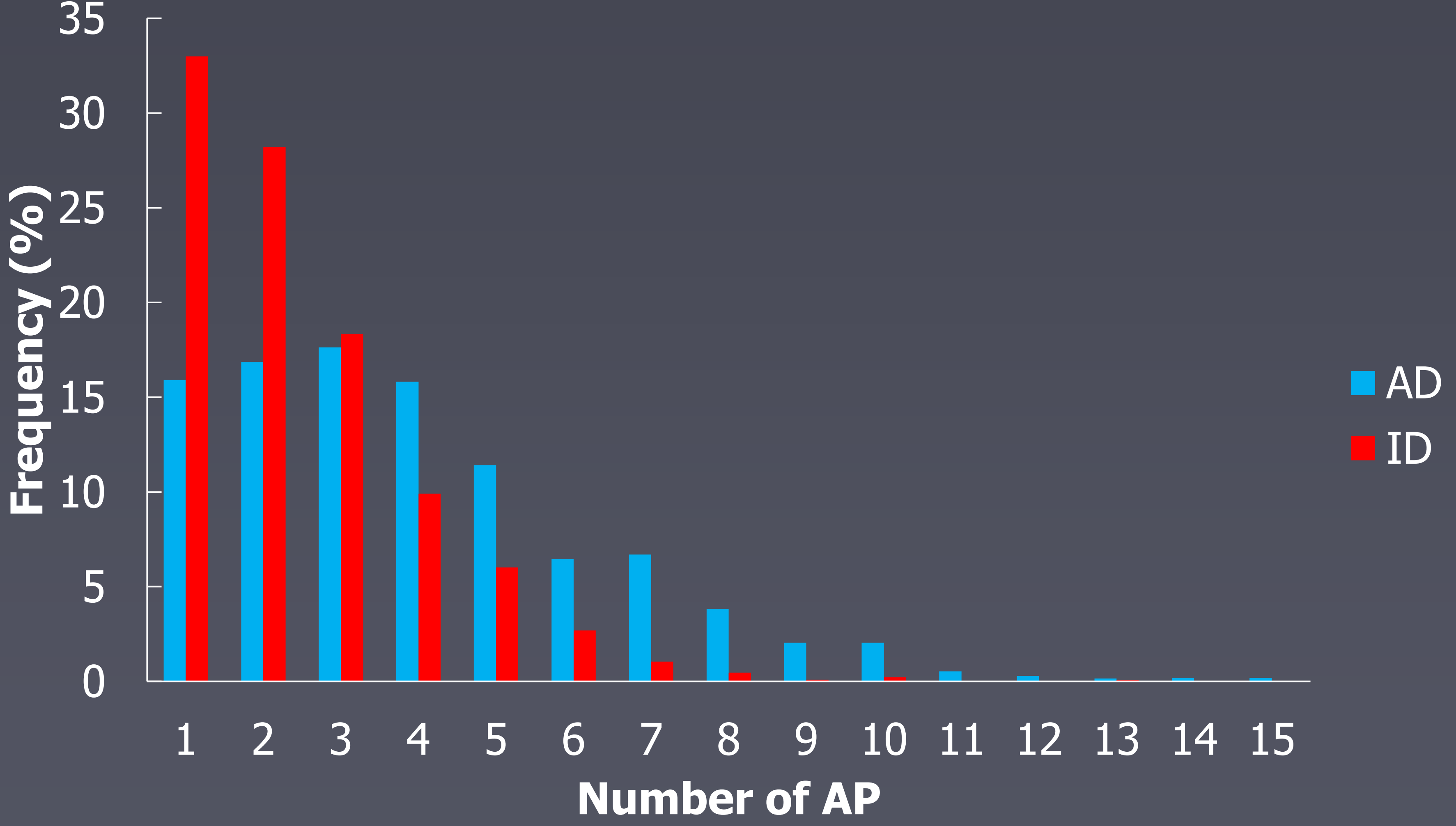
When syllables with BPM are removed

- Body of the utterances
  - “Max, min, mean: AD < ID
  - Range (st): AD > ID

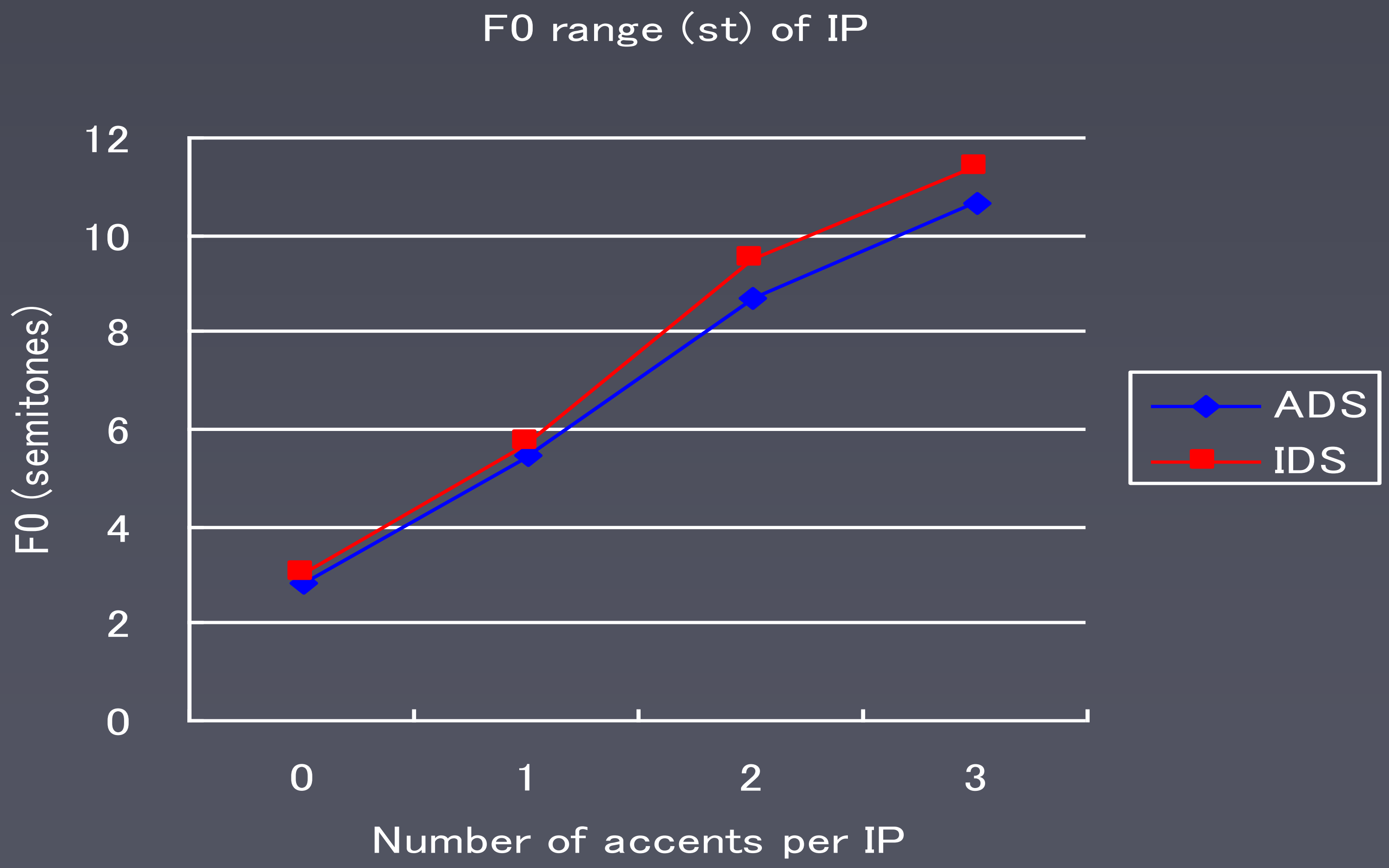




Frequency of the number of AP in an utterance



When IP is long & contain more accented words, pitch expanses



Pitch exaggeration in Japanese IDS

- No pitch exaggeration in overall utterances.
  - (Replication of Fernald et al, 1989)
- But;
- Pitch expansion found in syllables with BPM
  - This expansion *DOES* sound *exaggerated!*
- Larger pitch range in ADS, except for BPM
- ADS utterances are generally longer, contain more accented words, and have larger ranges.
  - This expansion does *NOT sound exaggerated*!

Intonation of Japanese IDS

- Exaggeration of intonation is an example of dynamic nature of prosody.
- Speakers' desire to exaggerate intonation maybe universal…
- But how to do so is constrained by the prosodic structure of the language.
- Examination of IDS register allowed us to examine this dynamic property.

Outline

1. RIKEN Japanese mother-infant conversation corpus (R-JMICC)
2. Exaggeration of intonation (Igarashi et al., JASA, 2013)
3. Realization of phonological rule (Martin, et al., Cognition, 2014)

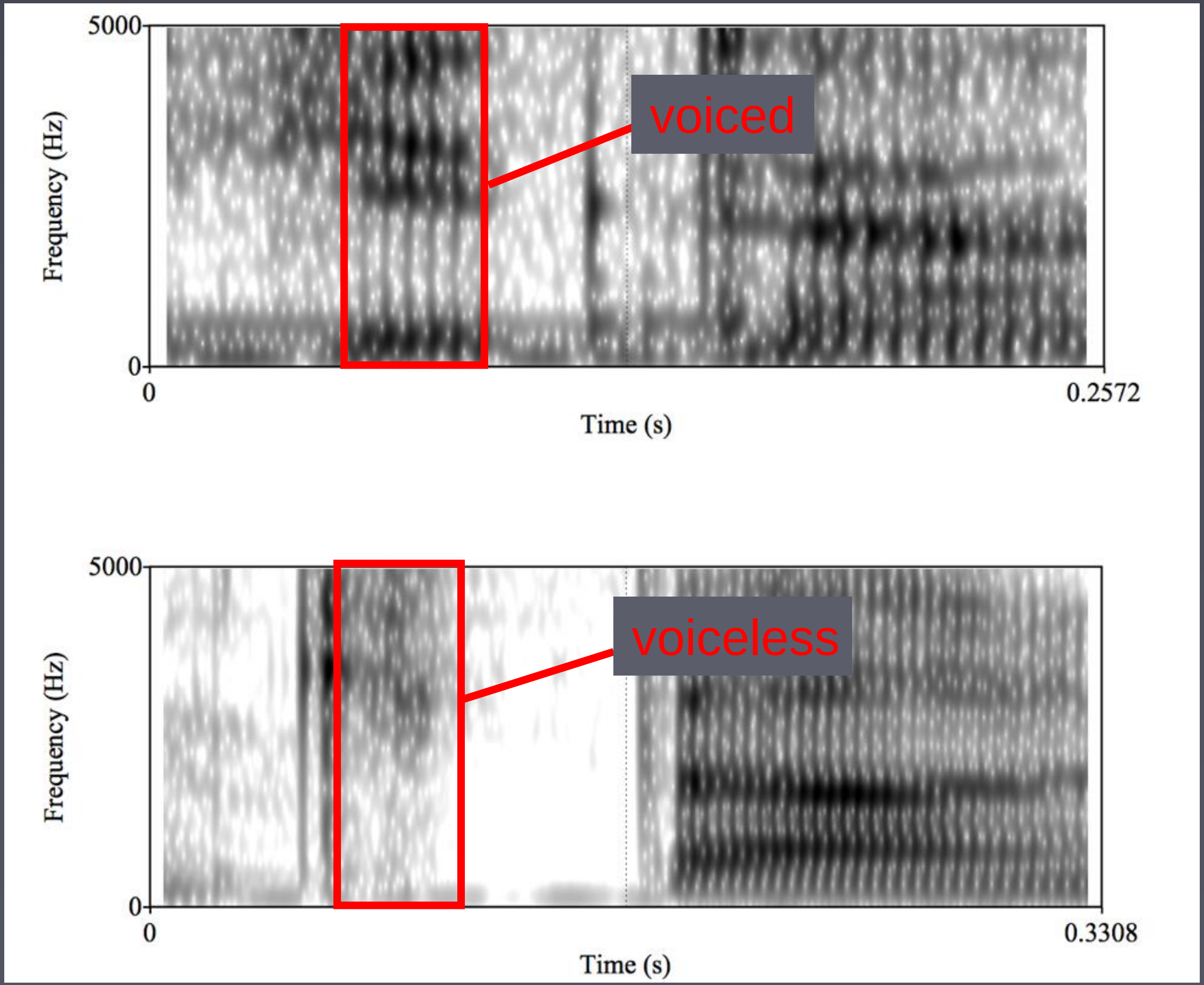
Phonological Rule: Vowel devoicing

Martin, Utsugi, & Mazuka, Cognition, 2014

- Rule:
  - Japanese high vowels /i/ & /u/([ɯ]) are devoiced between voiceless consonants, or following voiceless consonants word finally.
  - "kiki" (emergency) vs "kaki" (oyster)

High vowel devoicing

- Two tokens of *kita* ‘came’



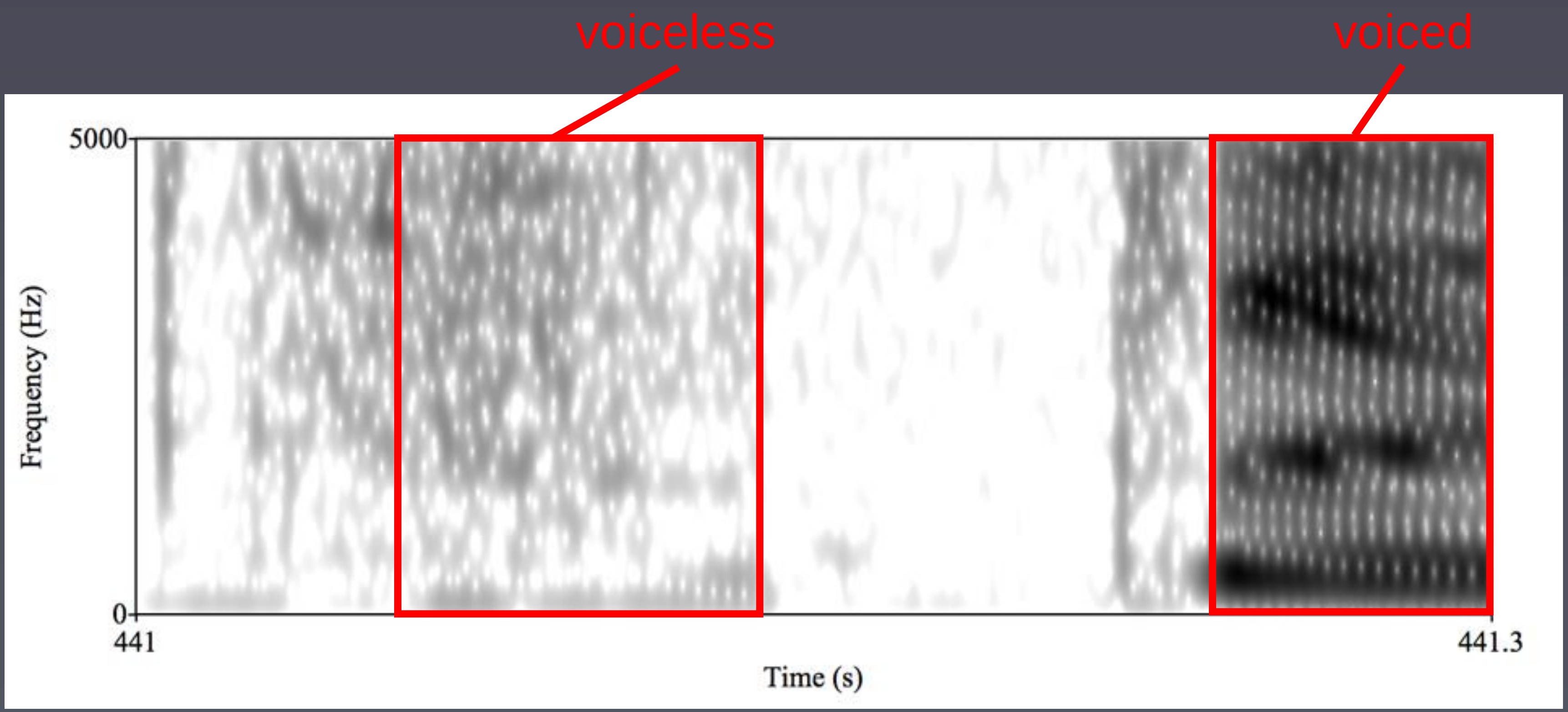
Phonological rule: Vowel devoicing

Martin, Utsugi, & Mazuka, Cognition, 2014

- Rule:
  - In Japanese, high vowels /i/ & /u/([ɯ]) tend to be devoiced between voiceless consonants, or following voiceless consonants word finally.
  - "kiki" (emergency) vs "kaki" (oyster)
- But
  - Non-high vowels can be devoiced also

Non-high vowel devoicing

- Devoiced /o/ in *soto* ‘outside’



Devoicing does not occur 100%

1. Devoicing of high vowels occur at much higher rate than non-high vowels
2. Devoicing rate changes by phonological contexts, speech rate, etc.
3. Speech register is a factor
   - Teachers devoice less when talking to hearing impaired children (Imaizumi, Hayashi, & Deguchi, 1995)

- *What determines devoicing rate?*

Data

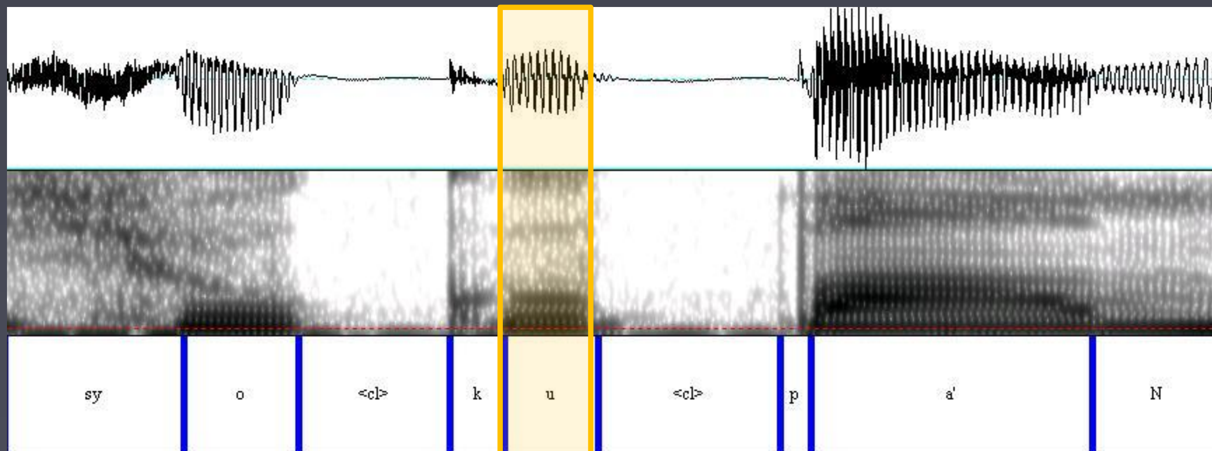
RIKEN Japanese Mother-Infant Conversation Corpus		
IDS	Playing with infant (average age 20 mo)	30 min each
ADS	Speaking with experimenter	10 min each
RS	Reading list of sentences (20 of 22 moms)	10 min each

Data

RIKEN Japanese Mother-Infant Conversation Corpus		
IDS	Playing with infant (average age 20 mo)	30 min each
ADS	Speaking with experimenter	10 min each
RS	Reading list of sentences (20 of 22 moms)	10 min each

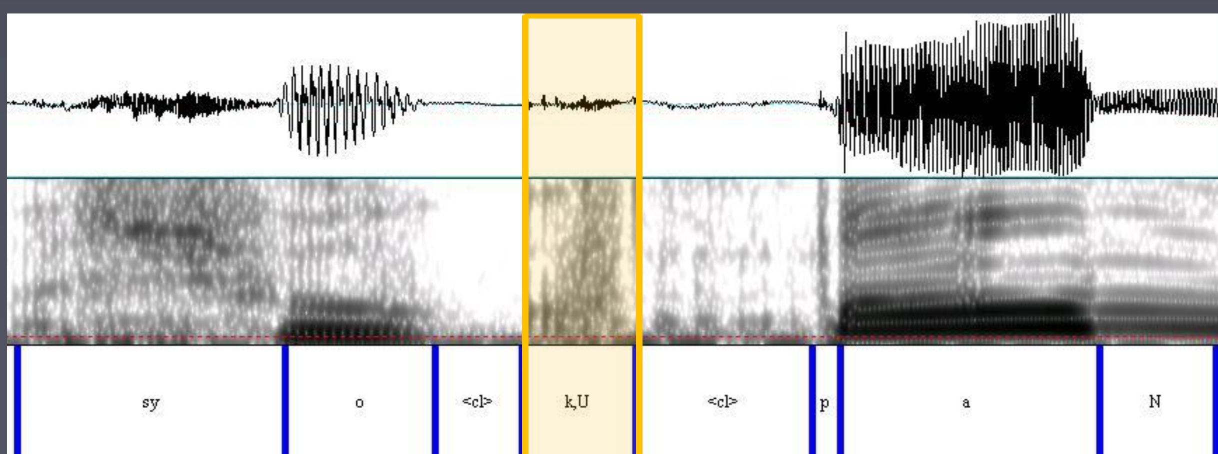
- Every vowel occurring between two voiceless consonants was labeled voiced or voiceless

Segmental labels



'bread' without devoicing

sy o <cl> k **u** <cl> p a N



'bread' with devoicing

sy o <cl> k **U** <cl> p a N

Data

RIKEN Japanese Mother-Infant Conversation Corpus		
IDS	Playing with infant (average age 20 mo)	30 min each
ADS	Speaking with experimenter	10 min each
RS	Reading list of sentences (20 of 22 moms)	10 min each

- Every vowel occurring between two voiceless consonants was labeled voiced or voiceless

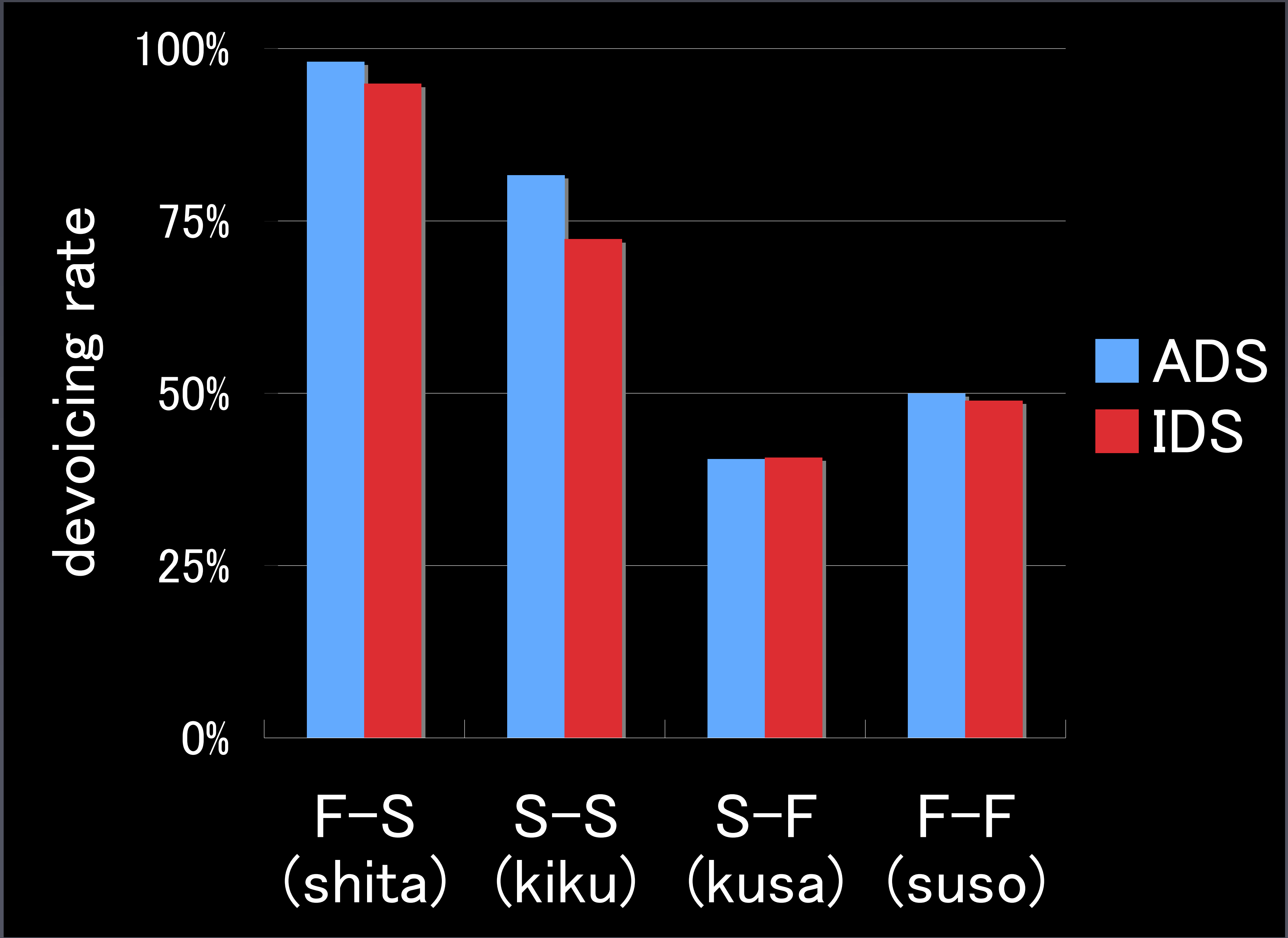
Devoicing rate

$$\frac{\text{number of devoiced vowels}}{\text{number of vowels between voiceless consonants}}$$

Linear Mixed Effects Models

- Dependent variable: voicing
- Fixed factors:
  - Speech style (RS, IDS, or ADS)
  - Vowel
  - Speech rate (moras per second)
  - Breathiness (mean H1-H2 of utterance)
  - Preceding context (affricate, fricative, stop)
  - Following context (affricate, fricative, stop)
  - Preceding * following
- Random factors:
  - Speaker
  - Word

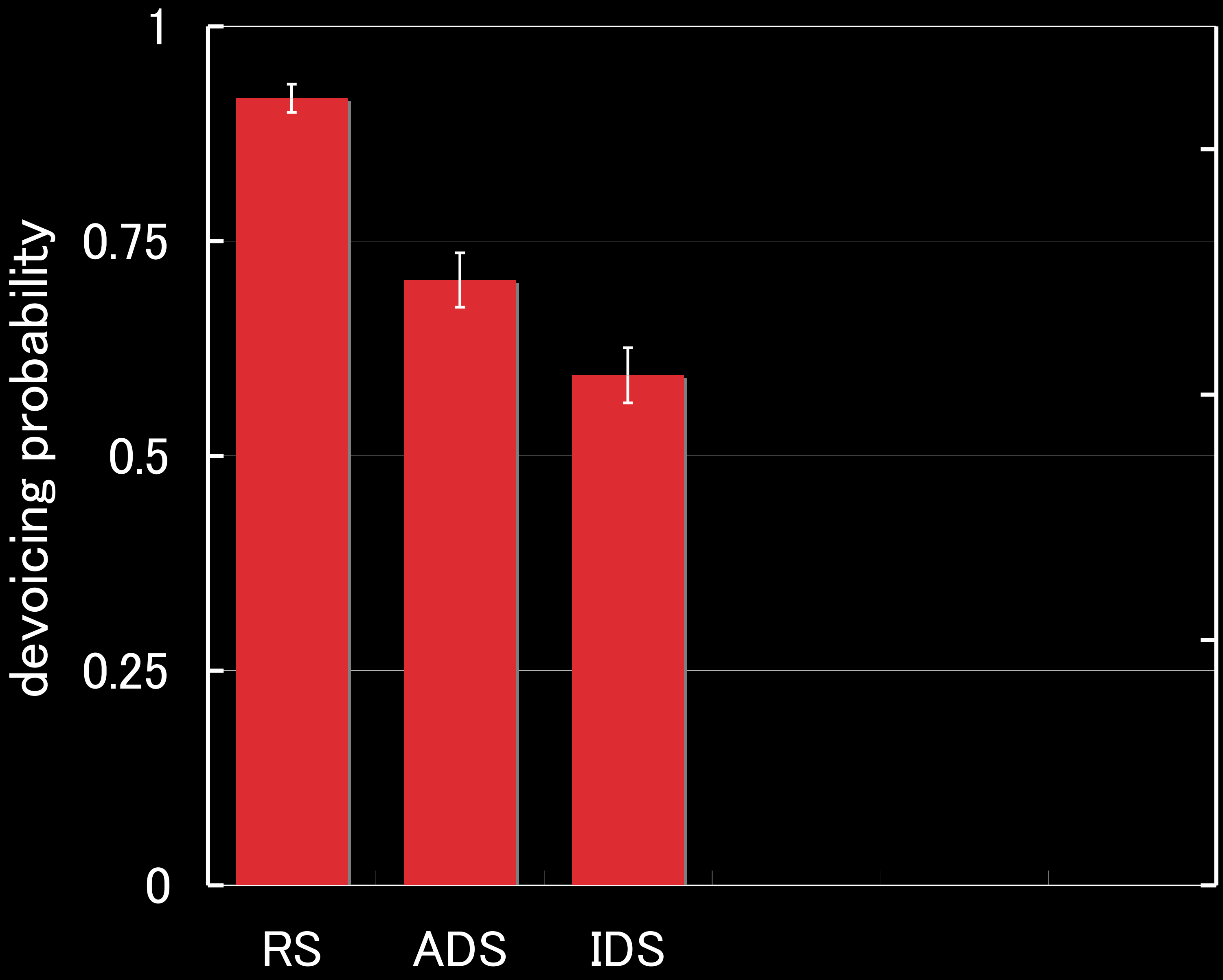
Segmental Contexts Matter



Linear Model Results

- Compared two models for each vowel height: one with SPEECH STYLE predictor, one without
- Including SPEECH STYLE significantly improves the fit of the model
  - High vowels: $\chi^2(1) = 37.4$, $p < 0.001$
  - Non-high vowels: $\chi^2(1) = 32.2$, $p < 0.001$
- Devoicing rate differences across speech styles are not just the result of subsidiary factors

Comparing All Speech Styles



Possible Explanations 1

1. Need to learn a phonological rule
   - Mothers want to provide more consistent input to infants

- Prediction:

  IDS > ADS/RS at least for high vowels

- Results

  ADS > IDS. Not supported by data.

Possible Explanations 2

2. Speakers want to increase intelligibility
   - Devoiced vowels are less intelligible (Gordon, 1988)
   - Adults find it difficult to identify devoiced vowels presented in nonce words (Beckman & Shoji, 1984; Cutler, et al., 2009)

► Prediction: ADS > IDS/RS

High vowels = Non-high vowels

► Results:

High vowels and non-high vowels showed an opposite pattern.

Why do high and non-high vowels behave differently?

- Two types of mechanism
  - Phonological: Speakers intend to devoice vowel, do so at abstract symbolic level (Tsuchida 1997, 2001; Teshigawara, 2002; Varden, 1998, 2010)
  - Phonetic: Speakers intend to voice vowel, but devoice inadvertently through gestural overlap (Imaizumi, Hayashi, & Deguchi, 1995; Jun & Beckman, 1993)
- Both mechanisms may operate in different dialects of Japanese (Fujimoto, et al, 1998; Fujimoto, 2004)

Why in opposite pattern?

- High vowels
  - Adults find it easier to recognize actual words with devoiced vowels (Ogasawara and Warner, 2009)
  - For adults, they are used to hearing words with high vowels devoiced. Devoicing known words -> better intelligibility.
  - For infants, most words are unknown. Devoicing -> worse intelligibility
  - The same goal (increasing intelligibility) results in the opposite results for adults & infants.

Why in opposite pattern?

- Non-high vowels
  - Even adults are not used to hearing words with non-high vowels devoiced.
  - Articulatory factors such as speech rate and breathiness of voice may account for at least part of the difference.

Discussion

- Devoicing rate dynamically changes in three speech registers.
- An opposite pattern of devoicing rate changes were observed in high- and non-high vowels
- Different factors that impact the two types of devoicing were revealed by examining distinct speech registers

Conclusion

- Phonological phenomena have dynamic aspects.
- They are difficult to capture in laboratory recordings.
- Comparisons of speech registers whose function is relatively easily defined can provide an window into dynamical aspects of real speech.

Thank You

LabPhon 14 - The 14th Conference on Laboratory Phonology

Tokyo, July 25-27, 2014

Important Dates

Conference Dates:

- **July 25-27, 2014**

Possible Dates for Satellite Workshops:

- **July 24, 2014**
- **July 28, 2014**

Abstract and Proposal Submission:

- **Abstract submission period: December 1, 2013 - January 31, 2014**
- **Notification of acceptance: March 15, 2014**

Early Registration:

- **April 1 - May 31, 2014**

10-2 Midori-cho, Tachikawa City, Tokyo, Japan 190-8561

LabPhon 14 - The 14th Conference on Laboratory Phonology

Tokyo, July 25-27, 2014

Call for Papers

- The Call for Papers has been closed.

Oral Presentations and Poster Presentations

- Abstracts are solicited for contributed papers for presentation as 20-minute oral contributions or as posters. Contributions relating to the conference themes are especially encouraged; there will also be sessions for non-thematic papers.
  - Abstract submission period: December 1, 2013 - January 31, 2014 (Honolulu time, UTC-10)
  - Notification of acceptance: March 15, 2014

Thematic and Non-thematic Sessions

- The overall theme for the conference is "Laboratory Phonology beyond the Laboratory: Quantitative Analyses of Speech Produced outside the Phonetics Laboratory". Our goal is to bring together researchers from phonology, phonetics, and adjacent psycho- and neurosciences and to seek to advance these disciplines by encouraging the joint pursuit of interdisciplinary research questions. Specific topics that address this theme are the following:
  - *fieldwork-based studies, particularly on endangered languages/dialects*
  - *corpus-based approaches to spontaneous speech*
  - *acquisition of L1 phonology/prosody*
- Non-thematic sessions (both oral and poster) will include contributions to other topics of interest to the LabPhon community.

Financial Support

- We plan to provide financial support for at least 30 student presenters (either oral or poster presentations). If you are a student and the sole or first author of a presentation, you are eligible. The money available will cover lodging expenses for up to four nights in Tokyo. We will not be able to help with transportation expenses.
- Shortly after the abstract review process is completed, we will send more detailed information to the authors of all accepted abstracts.

Abstract Submission

- To submit an abstract, attach both the abstract itself (as both a Word and a PDF file) and a filled-out submission form (see below) to an e-mail message addressed to:

  labphon14@ninjal.ac.jp

- LabPhon 14
- Guest Speakers
- Important Dates
- Call for Papers
- Program
- Registration
- Information for Presenters
- Information for Participants
- Accommodation and Travel
- Venue & Maps
- Committees

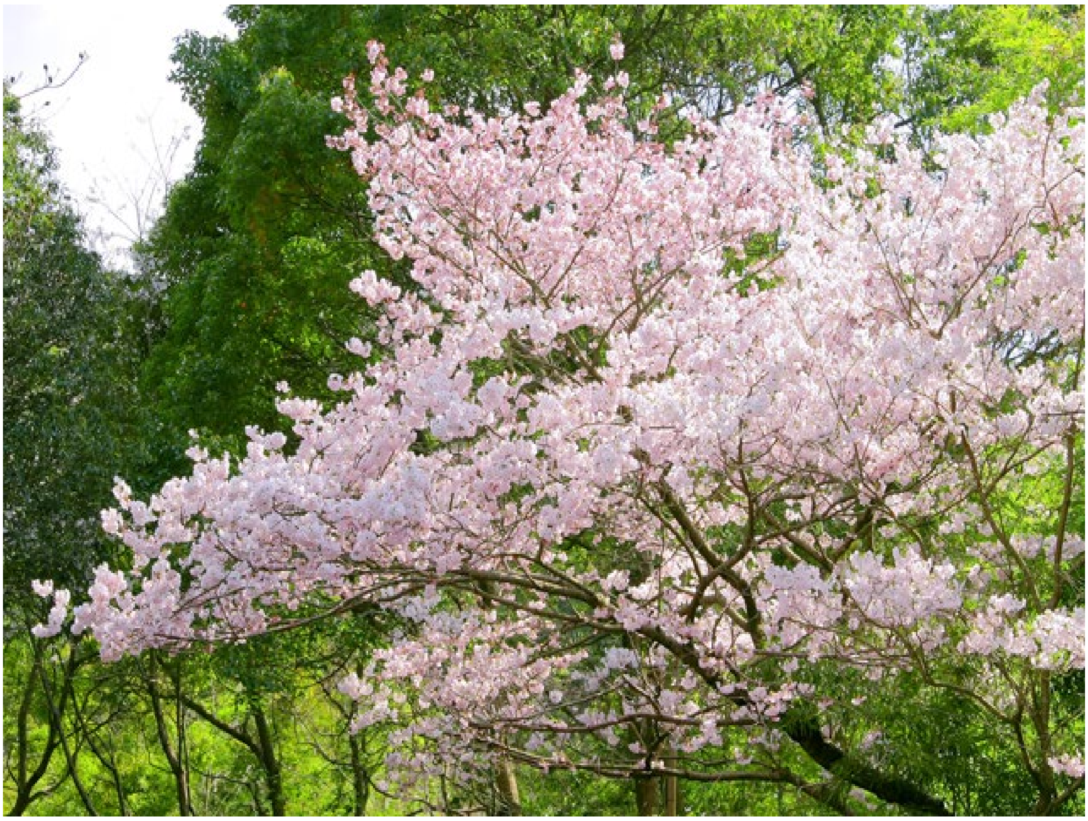
- The maximum length for the abstract is one page. This length restriction will be strictly enforced. A Word document template will be made available on this page soon. Please make sure that all special fonts and symbols are embedded in the PDF document and display properly.
- In addition, a submission form (see below) must accompany each abstract, attached it to the same e-mail message. The submission form can be either a Word file or a PDF file.
- One individual may be an author on no more than two submitted abstracts. All abstracts must be written in English. Presentations will be given in English. The abstracts will be evaluated anonymously by the scientific committee.
- The submission form asks you to identify which of the conference themes your paper relates to. If your paper does not relate to any of the themes, please select “non-thematic”. The submission form also asks you to specify whether you want to have your abstract considered for (1) either oral or poster presentation, (2) oral presentation only, or (3) poster presentation only. Please specify one of these three categories.

Abstract Template and Submission Form

- Abstract template
- Submission form

Publication of a Special Issue of Laboratory Phonology

- One issue of Laboratory Phonology is reserved for papers from LabPhon 14 (guest editors Haruo Kubozono, Kikuo Maekawa, and Timothy J. Vance). Owing to space limitations, the special issue will be restricted to papers that address the main theme of the conference "Laboratory Phonology beyond the Laboratory: Quantitative Analyses of Speech Produced outside the Phonetics Laboratory". Authors of papers in this category will be contacted personally. All authors whose papers are not considered for the special issue are encouraged to submit their manuscripts to the Laboratory Phonology journal as regular submissions.



Back to home Back Top

Sitemap

LabPhon 14 - The 14th Conference on Laboratory Phonology

Tokyo, July 25-27, 2014

Program

Conference Schedule

Thursday, July 24	Friday, July 25	Saturday, July 26	Sunday, July 27	Monday, July 28
	8:30-9:20 Registration			
	9:20-9:30 Welcome			
	9:30-10:50 ORAL SESSION 1	9:10-10:30 ORAL SESSION 5	9:10-10:30 ORAL SESSION 9	Satellite Workshop: MORNING SESSION
	Coffee Break	Coffee Break	Coffee Break	
	11:10-12:10 ORAL SESSION 2	10:50-12:10 ORAL SESSION 6	10:50-12:10 ORAL SESSION 10	
	Lunch	Lunch (LP Journal meeting)	Lunch	Lunch
13:30-17:30 LABPHON 14 Registration	13:10-14:30 ORAL SESSION 3	13:10-14:30 ORAL SESSION 7	13:10-15:00 POSTER SESSION 3 (with coffee)	
14:30-17:30 COLLOQUIUM - D. Robert Ladd - Anne Cutler	14:30-16:20 POSTER SESSION 1 (with coffee)	14:30-16:20 POSTER SESSION 2 (with coffee)	15:10-16:10 ORAL SESSION 11	Satellite Workshop: AFTERNOON SESSION
17:30~ Wine reception	16:30-18:10 ORAL SESSION 4	16:30-17:50 ORAL SESSION 8	16:10-16:30 GENERAL DISCUSSION	
			16:30-16:35 Farewell	
	18:10-18:40 ALP General Meeting	18:00-20:00 Banquet and Fireworks		~ 17:30

The Conference Schedule can be downloaded **here**.

Abstracts of Presentations

The **Conference Handbook** (including the **abstracts**) can be downloaded **here**.

Oral Presentations

The program of the **oral** sessions, can be downloaded **here**.

Poster Presentations

The program of the **poster** sessions can be downloaded **here**.

LabPhon 14 - The 14th Conference on Laboratory Phonology

Tokyo, July 25-27, 2014

Registration

Registration has closed

- **Registration closed on May 31, 2014 (midnight, Japan time).**
- **The interest in LabPhon14 has exceeded all expectations. As a result, the number of participants who have registered during the early registration period will fill the conference venue to capacity.**
- **Consequently, there will be NO late registration or on-site registration.**

Early Registration

- **April 1 - May 31, 2014**

▪ Registration will be possible from April 1, 2014 on.
▪ Payment must be made by credit card (VISA or MasterCard).
▪ Registrants will receive an official receipt for their registration fees at the registration desk in July unless they request otherwise.

- **Click here for online registration (from April 1 on)**

Registration Fees

- **Registration fees cover admission to the conference, conference kit, conference dinner, and coffee breaks. The fees for non-members also include a one-year membership fee for Laboratory Phonology (US$80 for students and US$125 for non-students), which, in turn, includes a subscription to the journal.**
- **LabPhon members must have completed the payment of their annual membership fee for 2014 directly to the Association if they wish to register as members.**
- **Those who register as a student are required to send their student ID (PDF) to labphon14@ninjal.ac.jp immediately after completing the online registration.**
- **So as not to exceed the capacity of the conference venue, registration will be limited to 200 participants. Registration is first come first served, so please register as soon as possible.**

Registration as:	Early Registration (before June 1)	Late Registration (before July 15)	On-Site Registration (Cash payment only)
LabPhon Member	JP¥6,500 (about US$65)	JP¥6,500	JP¥9,000
Non-Member	JP¥19,000 (about US$190)	JP¥22,000	JP¥27,000
Student LabPhon Member	JP¥1,500 (about US$15)	JP¥1,500	JP¥3,000
Student Non-Member	JP¥9,500 (about US$95)	JP¥12,000	JP¥15,000

- **Participants who register in advance (either early or late) will also receive lunch each day of the conference. Unfortunately, lunches cannot be provided for participants who register on-site, since the lunches must be ordered several days in advance.**
- **Cancellation policy: The Japanese financial system makes refunds very costly. For written cancellations received before July 15, 2014, we will refund the conference fee less JPY5,000 (about US$50) handling charges. After July 15 no refund is possible.**

Visa Information for Participants from Abroad

- **The regulations for entering Japan as a visitor differ significantly depending on the country that issued your passport. For many countries, no visa is required to enter Japan as a tourist and stay up to 90 days. However, for many other countries, a visa is required even for a short stay. If you are not sure about whether you need a visa, please contact your nearest Japanese embassy or consulate.**

10-2 Midori-cho, Tachikawa City, Tokyo, Japan 190-8561

LabPhon 14 - The 14th Conference on Laboratory Phonology

Tokyo, July 25-27, 2014

Information for Presenters

- This PDF file contains the information given on this page (version: 2014.06.01).

General Information

- The conference website is available at the following URL. Please check it for the updated program and other information. https://www2.ninjal.ac.jp/past-events/labphon14/
- The official language of the conference is English.
- We are expecting a total of 200 participants at the conference.
- For all presentations, at least one of the authors must pre-register (with payment) by May 31st.
- A pre-conference colloquium is scheduled on the afternoon of July 24 with a reception following.
- Also, the conference party and fireworks will be held at the conference venue on the evening of July 26, between 18:00 and 20:00. The fee for these events is included in the registration fee. We hope that you will take these opportunities to get to know other participants.
- Because there are fireworks in the park near NINJAL on the 2nd day of the conference, we recommend making your hotel reservation as early as possible. For any inquiries, please contact us at [labphon14@ninjal.ac.jp].

Specific Instructions for Speakers

- Each talk is 20 minutes long including the time for questions and comments. Please aim to speak for up to 15 minutes and leave at least 5 minutes for questions and comments.
- A projector and speakers will be available at the conference. If you are doing a PowerPoint presentation, it would be best if you use your own laptop, as there may be problems (with alignment, fonts, etc.) if you switch computers, but if you wish, we would be happy to lend you a laptop. If this is your preference, please let us know in advance whether you need PC/Windows or Mac.
- If you have a handout for your talk, please make 200 copies and bring them to the registration desk at the beginning of the morning session of your presentation day.

Specific Instructions for Poster Presenters

- Your poster should be in English and maximally A0 size, i.e., 118.9 cm (wide) by 84.1 cm (high).
- You will have a poster board this size to display your poster (more about this below). Posters can be affixed to the board with pins. We will prepare sufficient pins for each poster.
- Please try to use a large font size in your poster, so that the audience can easily understand your poster when they stand, say, 1 meter away from it.
- The poster session will be held in the space on the second floor just outside the Lecture Hall where the oral presentations will be given, and also on the first floor. Poster presentations are divided into three groups, Poster session 1 (Day 1), Poster session 2 (Day 2) and Poster session 3 (Day 3), and numbered sequentially in each group (e.g., “P1-15”). Please check your poster number in the program

LabPhon 14
Guest Speakers
Important Dates
Call for Papers
Program
Registration
Information for Presenters
Information for Participants
Accommodation and Travel
Venue & Maps
Committees

in advance. Please affix your poster to the board by noon on the day of the presentation. Please make sure to use the board marked with your poster number.

- **You can keep your poster on the board until the end of the final session of the day.**
- **Please feel free to distribute subsidiary materials such as handouts, photocopies of your poster, and copies of your paper. You can also use your laptop computer if you find it helpful.**

Back to home   Back   Top

Sitemap

10-2 Midori-cho, Tachikawa City, Tokyo, Japan 190-8561

LabPhon 14 - The 14th Conference on Laboratory Phonology

Tokyo, July 25-27, 2014

Information for Participants

This PDF file contains the information given on this page.

Information for LabPhon 14 Participants

Please visit the conference website (https://www2.ninjal.ac.jp/past-events/labphon14/) from time to time as it is regularly updated. It contains many pieces of useful information including an updated program of the main conference (July 25-27) and its satellite events (July 24 and 28). It also has useful information about hotels, journey to the conference venue, tourist information, etc. Below we give some information which you might find particularly useful.

Hotels

As already noted on the conference website, hotels near the conference venue are already very crowded. Most hotels in Tachikawa are already fully booked on Saturday, July 26, because of the annual Fireworks Festival in Tachikawa which attracts over 300,000 people every year. If you haven't booked your hotel yet, we recommend that you do so as soon as you can. The conference website lists more than ten hotels near the conference venue, including three big hotels that offer discount rates for LabPhon 14 participants. If you need any assistance, please contact the organizers at labphon14@ninjal.ac.jp.

Lunch and dinner for non-participants

If someone accompanies you to the conference, you can book a conference dinner (Saturday, July 26) for him/her at JPY 3,000 per person (children under 12 are free). Please contact us before July 15 if you decide to bring someone with you to this banquet. In the second half of the banquet, you can enjoy the firework festival from the balcony of the building. Please note, however, that we cannot provide lunch for accompanying persons; please bring lunch for them.

Satellite events

There will be a pre-conference colloquium featuring Bob Ladd and Anne Cutler on Thursday, July 24. This will be followed by registration (name tag, conference booklet, etc.) and a complimentary wine reception. There is no need to make a booking for these events. On the other hand, the satellite workshop on Monday, July 28, requires pre-registration. If you are interested in attending this workshop, please visit the following website and register for it: http://phonetik.uni-muenchen.de/labphon14-satellite/registration.html)

Journey to NINJAL in Tachikawa

For those who are not familiar with Tokyo, the conference website gives details about the journey from Narita and Haneda Airports to Tachikawa and to NINJAL. Tourist information is available, too, including the links to various useful websites. In addition, complimentary guidebooks about tourists' spots in Tokyo will be available at the registration desk, in several languages including English and Japanese.

Wifi and PC at NINJAL

Wifi is available free of charge at NINJAL. The password can be found in the conference booklet you will receive at the registration desk. A PC is available for oral presentations upon request. If you want to use this service, please contact us soon.

Restaurants near the conference venue

There are hundreds of restaurants and shops near JR Tachikawa Station, which is 20 minutes on foot from the conference venue. Most restaurants are reasonable in price, including those on the top floor of the two big department stores near the station, Takashimaya and Isetan. At the conference, we plan to prepare a dinner map for vegetarians.

The Vowel

The Phonetic Society of Japan is selling the last 30 copies of The Vowel: Its nature and Structure by Tsutomu Chiba and Masato Kajiyama. This pioneering book is out of print and costs $195 through Amazon. This book is available at the registration desk at the special conference rate of JPY1,000 (ca $10) on a first-come first-served basis.

If you have any questions or need any assistance, please contact us at labphon14@ninjal.ac.jp.

We look forward to welcoming you to Tokyo in July.

LabPhon 14 Organizers

10-2 Midori-cho, Tachikawa City, Tokyo, Japan 190-8561

LabPhon 14 - The 14th Conference on Laboratory Phonology

Tokyo, July 25-27, 2014

Accommodation and Travel

Acommodation around Tachikawa

- We recommend that you make your hotel reservations well in advance (at least one month before the conference if possible). The Palace Hotel Tachikawa and the Tachikawa Grand Hotel are particularly convenient because they are both stops for the airport bus to and from Narita.

Hotel Discount

- The Palace Hotel Tachikawa is offering a special rate for LabPhon14 participants (Y9,700 per night for a single room). Please download this Word document for more information.
- Tachikawa Grand Hotel is offering special rates for LabPhon14 participants (both single and twin rooms). Please download this Word document for more information.
- Tachikawa Washington Hotel is offering special rates for LabPhon14 participants (single, twin and double rooms). Please download this Word document for more information.

Links to hotels within walking distance of NINJAL

Price guide:
: most expensive
: relatively less expensive
: least expensive

- **Palace Hotel Tachikawa**

http://www.palace-t.co.jp/english/index.html
http://www.palace-t.co.jp/index.html

- **Hotel Mets Tachikawa**

http://www.jrhotelgroup.com/eng/code/codeeng160.htm
http://www.hotelmets.jp/tachikawa/

- **Tachikawa Grand Hotel**

http://tachikawa.khgrp.co.jp/

- **Tachikawa Washington Hotel**

http://tachikawa.washington-hotels.jp/
http://www.tachikawa-wh.com/

- **Toyoko Inn Tachikawa-eki Kitaguchi**

http://www.toyoko-inn.com/hotel/00245/index.html

- **Showa Hotel**

http://www.showa-hotel.jp/

- **Tachikawa Hotel**

http://www.tachikawa-h.jp/index.html

- **Business Hotel Homare**

http://hotelhomare.com/

Royal Authentic Hotel Tachikawa

http://www.authentic-hotel.co.jp/

The Crest Hotel Tachikawa

http://www.cresthotel.co.jp/tachikawa/

Hotel Rex Tachikawa

http://www.rex-tachikawa.jp/

Tachikawa Regent Hotel

http://www.tachikawa-regent-hotel.jp/

Super Hotel Tokyo-JR Tachikawa Kitaguchi

http://www.superhotel.co.jp/s_hotels/tachikawakita/tachikawakita.html

Tourist Information

Japan

http://www.att-japan.net/
http://www.japan-guide.com/
http://travel.rakuten.com/

Tokyo

http://www.gotokyo.org/en/index.html
http://guide.enjoytokyo.jp/en/

Tachikawa

http://www.tbt.gr.jp/english/index.html

Yokohama

http://www.welcome.city.yokohama.jp/eng/travel/

Kamakura

http://en.kamakura-info.jp
http://guide.city.kamakura.kanagawa.jp/

Hakone

http://www.hakone.or.jp/english/index.html

Nikko

http://www.nikko.or.jp/index.html
http://nikko.4-seasons.jp/

Kyoto

http://www.kyoto-magonote.jp/en/
http://www.kyoto.travel/

Osaka

http://www.osaka-info.jp/en/

10-2 Midori-cho, Tachikawa City, Tokyo, Japan 190-8561

LabPhon 14 - The 14th Conference on Laboratory Phonology

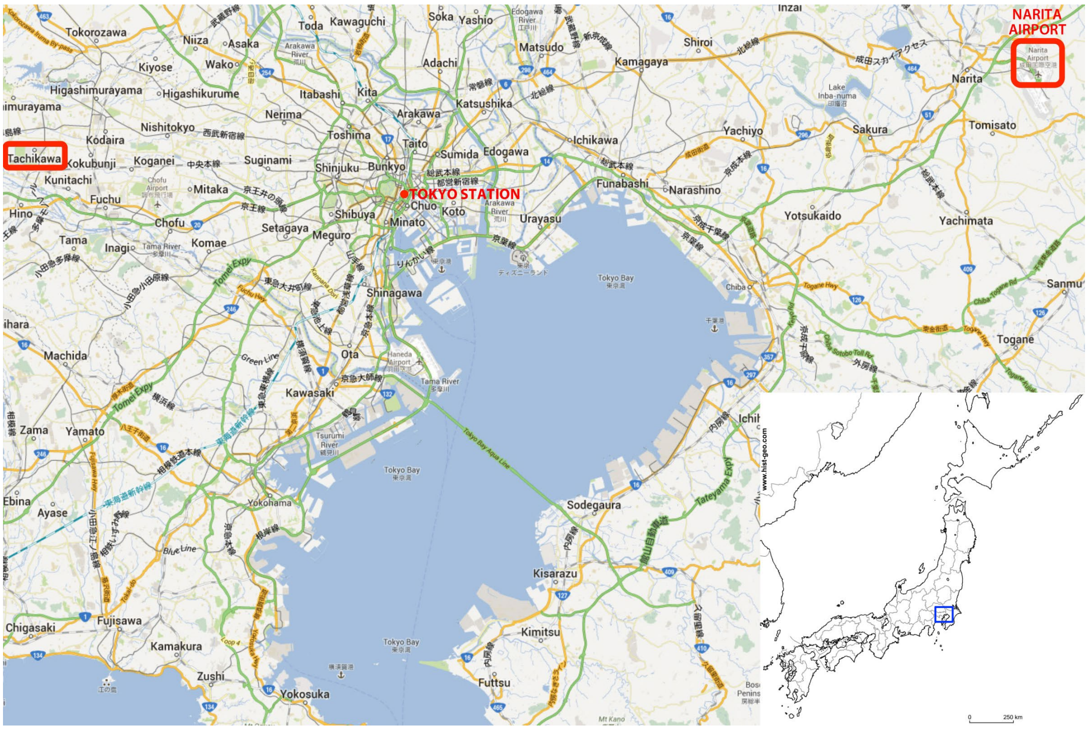
Tokyo, July 25-27, 2014

Venue & Maps

Getting to Tachikawa

Tokyo Airports

Most international flights to Tokyo arrive at Narita Airport (http://www.narita-airport.jp/en/), but there has recently been an increase in service to Haneda Airport (http://www.haneda-airport.jp/en/). NINJAL (the site of LabPhon 14) is located in the city of Tachikawa, which is about an hour by train from Tokyo Station.



Narita to Tachikawa

There are two options for getting from Narita Airport to Tachikawa.

By train

If you have purchased a Japan Rail Pass so that you can travel around by train before or after the conference, you can exchange your voucher for the actual pass at the JR East (East Japan Railway Company) station in the airport and take the Narita Express (http://www.jreast.co.jp/e/nex/narita.html).
There are two trains a day that will take you all the way to Tachikawa. They leave Narita Airport Terminal 1 at 19:45 and at 21:49. If you do not have a rail pass, this trip will cost Y3,980.
If you are not able to take one of these two trains, you will have to take the Narita Express from the airport to Shinjuku (http://www.jreast.co.jp/e/stations/e866.html). At Shinjuku, it will be necessary to transfer to an outbound Chūō Line commuter train. All outbound trains stop at Tachikawa. If you do not have a rail pass, this trip will cost Y3,550.

Please note:The Chūō Line is one of the most heavily used train lines in Japan, and the outbound trains are extremely crowded from mid afternoon until late at night. It will take at least 25 minutes to get from Shinjuku to Tachikawa, and you are very unlikely to find a seat. Also, it can be quite a challenge just to get on and off the train if you have a suitcase.

By bus

For most people, especially those with luggage, the Airport Limousine bus (http://www.limousinebus.co.jp/en/platform_searches/index/2/101#reserve) is a better option. The ticket counters are located just outside the exits from Customs and Immigration. There are 9 trips a day to Tachikawa, and the fare is Y3,500. The last departure each day is at 20:20–20:30 (depending on which of the three airport stops you board at). The trip typically takes about 2½ hours, but it varies considerably depending on traffic conditions. When you buy your ticket, the clerk will ask you where you want to get off. Choose either Tachikawa Station/Grand Hotel (the first stop) or Palace Hotel (the second stop), depending on where you are staying. Unless you are staying at the Palace Hotel, the first stop will probably be better, but be aware that it is actually quite a distance (about 300m) from Tachikawa Station.

Haneda to Tachikawa

There are two options for getting from Haneda Airport to Tachikawa.

By train

Unless you are familiar with the Tokyo commuter train network and do not have much luggage, we do not recommend traveling from Haneda to Tachikawa by train. The easiest route is to take the Tokyo Monorail (http://www.tokyo-monorail.co.jp/english/) from the airport to Hamamatsu-chō, the JR East Yamanote Line or Keihin-Tōhoku Line from Hamamatsu-chō to Tokyo Station, and the JR East Chūō Line from Tokyo Station to Tachikawa.

By bus

The Airport Limousine bus (http://hnd-bus.com/route/haijima.html) is the recommended option. The airport ticket counter is located just outside the exit from Customs and Immigration. There are 20 trips a day to Tachikawa, and the fare is Y1,500. The last departure each day is at 10 minutes past midnight. The trip typically takes a little less than 2 hours, but it varies considerably depending on traffic conditions. When you buy your ticket, the clerk will ask you where you want to get off. Choose either Tachikawa Railway Station (the first stop) or Palace Hotel (the second stop), depending on where you are staying. Unless you are staying at the Palace Hotel, the first stop will probably be better.

Finding your hotel

If this is your first trip to Japan, the area around a major railway station like Tachikawa may feel like a maze, especially if you arrive after dark. Your hotel's website should include a map that will help you to find your way, but if you get lost, almost anyone you approach will try to help you. Try to speak English slowly and clearly (although you might be lucky enough to run into someone who is highly proficient). People who work in convenience stores (7-Eleven, Lawson, Family Mart, Sunkus, etc.) usually know the area well. If necessary, you can ask for directions at the police box near the Krispy Kreme on the north side of Tachikawa Station (ground level) or at the information counter inside Tachikawa Station (located upstairs near the ticket gates).

Police box

Information center

Getting from Tachikawa Station to NINJAL

NINJAL

The National Institute for Japanese Language and Linguistics (NINJAL) is the host for LabPhon 14. NINJAL is located about 1.5km north of Tachikawa Station. You can get from Tachikawa Station to NINJAL on foot, by bus, by monorail, or by taxi.

On foot

It takes about 20 minutes by walk from Tachikawa station (take the North exit). The easiest route would be to take Nishi Odori St. (see the above map). [walking directions]

By bus

Buses run frequently between Tachikawa Station and NINJAL, and the ride takes about 5 minutes. There is a bus stop right in front of NINJAL. The bus is probably the best option on a rainy day, but it is not to easy to use if you do not know Japanese, so please follow the instructions provided. The fare is ¥190. [bus directions]

By monorail

The Tama Monorail stops both on the south side and on the north side of Tachikawa Station. The closest monorail station to NINJAL is Takamatsu, and the ride takes about 2 minutes from Tachikawa-kita (the Monorail station at the north side of Tachikawa Station). However, it takes about 7 minutes to walk from Takamatsu station to NINJAL. Please follow the instructions provided. The fare is ¥100. [monorail directions]

By taxi

There is a taxi stand on the on the north side of Tachikawa Station near the Stabucks on the gound level. The minimum fare of Y710 should be enough to get you from there to NINJAL. You can also catch a taxi on the south side of Tachikawa Station, but it will take quite a bit longer and cost more because the driver will have to take a circuitous route to get under the train tracks. A single taxi is big enough to carry 3 people comfortably, and it is usually possible squeeze in 4 people, with one passenger in the front seat. The driver pushes a button to automatically open the and close the left rear door (where passengers normally get in and out), and passengers are not supposed to open or close this door themselves. Most drivers do not speak English, but they will understand if you say “Receipt, please.” On the other hand, “NINJAL” will mean nothing to a taxi driver, so if you plan to take a taxi, you should print out the taxi directions provided so that you can show them to your driver. [taxi directions]

Back to home Back Top

Sitemap

10-2 Midori-cho, Tachikawa City, Tokyo, Japan 190-8561

LabPhon 14 - The 14th Conference on Laboratory Phonology

Tokyo, July 25-27, 2014

Committees

Local Organizing Committee

- Mafuyu Kitahara (Waseda University)
- Haruo Kubozono (Dept. of Linguistic Theory and Structure, NINJAL)
- Kikuo Maekawa (Dept. of Corpus Studies, NINJAL)
- Keiichi Tajima (Hosei University)
- Timothy Vance (Dept. of Linguistic Theory and Structure, NINJAL)
- Kiyoko Yoneyama (Daito Bunka University)

Executive Council of the Association for Laboratory Phonology

- D. Robert Ladd, President (2012-2014)
- Martine Grice, Vice-President (2012-2014)
- Matt Goldrick, Secretary (2012-2014)
- Caroline Smith, Treasurer (2012-2014)
- Grzegorz Dogil, Organizing committee LabPhon 13 (2010-2014)
- Haruo Kubozono, Organizing committee LabPhon 14 (2012-2016)
- Mary E. Beckman (2012-2014)
- Ioana Chitoran (2012-2016)
- Taehong Cho (2012-2016)
- José Ignacio Hualde (2012-2014)
- Jennifer Cole, Editor-in-Chief of Laboratory Phonology (2010-2015)

10-2 Midori-cho, Tachikawa City, Tokyo, Japan 190-8561

Leaky phonology and language design

Bob Ladd
University of Edinburgh
Tachikawa, 24 July 2014

Leaky grammar

- “Unfortunately, or luckily, no language is tyrannically consistent. All grammars leak.” (Sapir 1921, *Language*)

Leaky grammar

- “Unfortunately, or luckily, no language is tyrannically consistent. All grammars leak.” (Sapir 1921, *Language*)
- “Leakiness” is generally understood (descriptively) in terms of lexical exceptions and/or (descriptively and historically) in terms of competition between grammatical constraints and analogies.

Leaky phonology?

Morphophonology definitely leaks:

- E.g. Eng. trisyllabic laxing in *ins**a**ne/ins**a**nity; comp**e**te/comp**e**titive; f**i**nite/inf**i**nity;* etc. etc.

but not in *ob**e**se/ob**e**sity.*

Leaky phonology?

Morphophonology definitely leaks:

- E.g. Jap. *rendaku: nimai* 二枚 (にまい) + *shita* 舌 (した) ‘tongue’ ➔ *nimaijita* 二枚舌(にまいじた) ‘duplicity’;

but *kutsu* 靴 (くつ) ‘shoe’ + *shita* 下 (した) ‘below’ ➔ *kutsushita* 靴下 (くつした) ‘sock’. (**kutsujita* *くつじた)

Leaky phonology?

Otherwise, phonological theories are based on "non-leaky" assumptions:

- Specifiable finite inventory of phonemes
- Exhaustive analysis of words/morphemes
- Phonemes are meaningless
- No lexical exceptions (e.g. no unique phonemes, no exceptions to allophony)

Leaky phonology?

Summary of standard view:

- Phonology makes it possible to provide an exhaustive analysis of the shape of every word and morpheme of a language, and therefore ultimately of any utterance, ***independent of the grammatical analysis***.
- This is the idea of "double articulation" (Martinet) or "duality of patterning" (Hockett).

Duality of patterning

- Notion developed (independently?) by Martinet and Hockett in the 1940s and 1950s on the basis of Hjelmslev’s work.
- Two completely separate analyses of any utterance: as a string of ***meaningful*** elements (e.g. morphemes) (Hjelmslev’s “content plane”) and as a string of ***meaningless*** elements (e.g. phonemes) (Hjelmslev’s “expression plane”).

Duality of patterning

- Hockett saw duality of patterning as a key “design feature” of human language that distinguished it from other biological communication systems.
- This idea is influential in work on language evolution and animal communication, in particular birdsong (e.g. Yip 2006, Fitch and Jarvis 2013).

Duality of patterning

- Duality idea also helps drive assumption of a separate phonological structure for phrases, independent of syntactic structure. (Compare Kaisse 1985 with e.g. Selkirk 1984 or Nespor and Vogel 1986.)

Duality of patterning in signed language

- Development of Al-Sayyid Bedouin Sign Language (ABSL) (Sandler et al., e.g. 2011).

Duality of patterning in signed language

- Development of Al-Sayyid Bedouin Sign Language (ABSL) (Sandler et al., e.g. 2011).
- Isolated community affected by congenital deafness beginning in early 20th century. Now a significant proportion of community members are deaf.

Duality of patterning in signed language

- Development of Al-Sayyid Bedouin Sign Language (ABSL) (Sandler et al., e.g. 2011).
- Naturally developed sign language has arisen in 4-5 generations. Everyone signs, including hearing members (who speak Arabic).
- Development of language (esp. past 25 years) very well documented.

Duality of patterning in signed language

- Development of Al-Sayyid Bedouin Sign Language (ABSL) (Sandler et al., e.g. 2011).
- Signs ***evolve*** toward phonological analysability. Phonology ***(and therefore duality of patterning)*** "emerges" in new sign languages. (Cf. Frishberg 1975 on "iconicity" in American Sign Language.)

Duality of patterning in signed language

- Development of Al-Sayyid Bedouin Sign Language (ABSL) (Sandler et al., e.g. 2011).
- Phonology emerges ***after*** grammar and some lexicon are firmly in place. Related ideas about sign phonology put forth by Brentari and colleagues.
- Aronoff (2008): “In the beginning was the word”.

Duality of patterning in signed language

- Similar to Hockett’s conclusions about spoken language phonology – except that he was talking about language evolution in the species, not the development of individual languages.

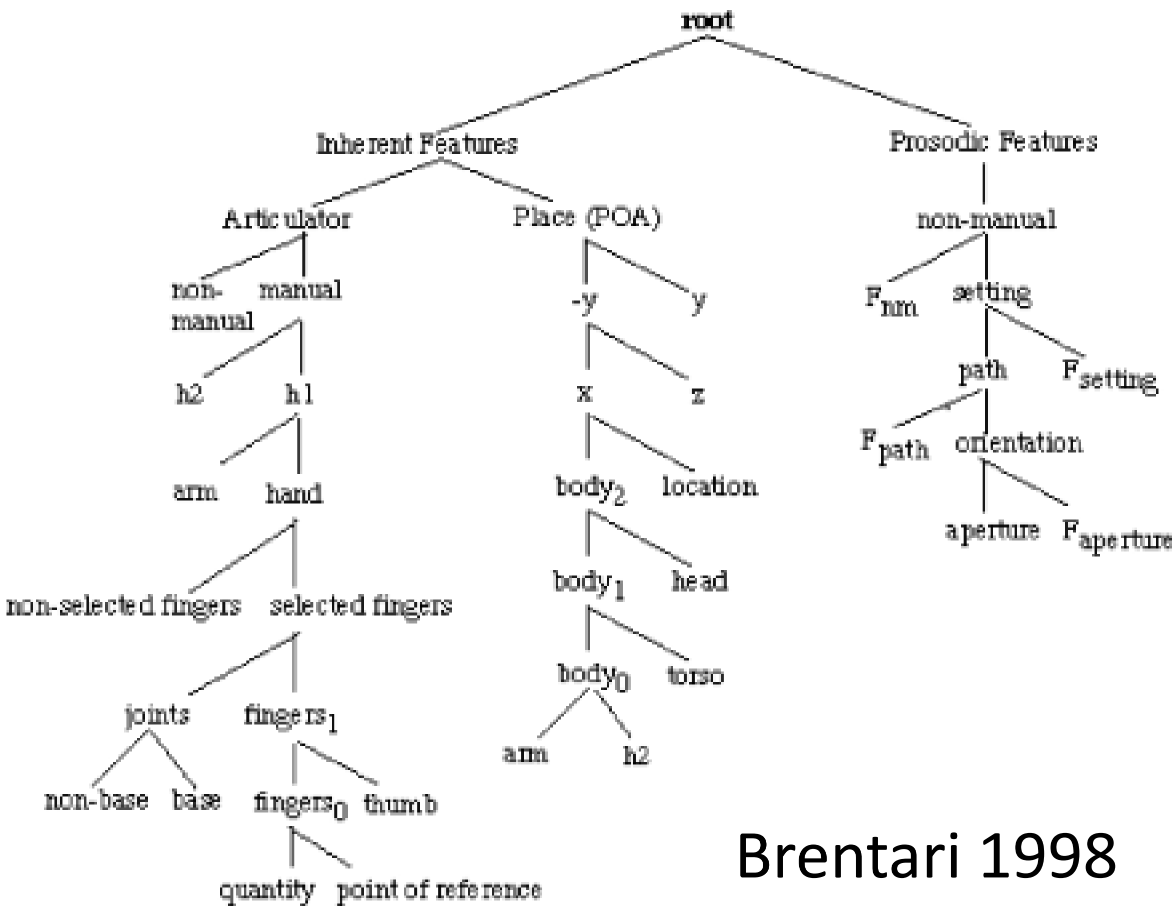
Duality of patterning in signed language

- Similar to Hockett's conclusions about spoken language phonology – except that he was talking about language evolution in the species, not the development of individual languages.
- ***Sign language work suggests a different understanding of duality of patterning.***

Duality of patterning in signed language

- Important problem in all sign language phonology is ***large number of primitives***.
- Classic analysis (Stokoe 1960) in terms of **handshape, location,** and **movement.**
- Elaborated since then; e.g. Brentari 1998

Duality of patterning in signed language



Brentari 1998

Duality of patterning in signed language

- ***Yet despite the large number of primitives*** it is sometimes difficult to analyse signs in terms of a fixed phonological inventory.

e.g. ASL “INTERNALIZE” is exceptional in having two places of articulation (chest and non-dominant hand). (Image : Brentari 1998)

Duality of patterning in signed language

- ***These problems are exactly what we should expect if phonology gradually emerges.*** Signs do not suddenly become exhaustively analysable phonologically. A residue of exceptions – “leaks” in the phonology – is therefore no surprise.

Duality of patterning in signed language

- ***These problems are exactly what we should expect if phonology gradually emerges.*** Signs do not suddenly become exhaustively analysable phonologically. A residue of exceptions – "leaks" in the phonology – is therefore no surprise.
- ***Are such leaks also found in spoken phonology?***

Leaky phonology in spoken language?

- Phonemes are not always meaningless:

(1) onomatopoetic or imitative words (e.g. Eng. *moo, bow-wow,* etc.; e.g. Jap. *giongo* 擬音語)

(2) “sound symbolism” of various types (e.g. Eng. *flash, flicker, flare, flame* etc.; Eng. <-*y*> in *sorry, holy, silly, happy* etc.; e.g. Jap. *gitaigo* 擬態語).

Leaky phonology in spoken language?

- Some words contain sounds that are not part of a language's normal phoneme inventory:

(1) Interjections, etc.

- e.g. Eng. *uh-huh, uh-uh*
- e.g. Eng. *ow, ouch*; Fr. *aïe*; Ger. *aua*; It. *aia*; Jap. /ite/ 痛っ; Eng. *yuck*; Fr. *beurk*; Ger. *igitt*; It. *peh*; Jap. /ge/ げっ

(2) Foreign words (e.g. Eng. *Debussy, loch*)

(3) "Special" words (e.g. Arabic *Allah*)

Leaky phonology in spoken language?

- Some allophonic rules have exceptions:

(1) Foreign words:

- e.g. Jap. /t/ before /i/
  tisshu ティッシュ 'tissue (paper)'
  charitii チャリティー 'charity'

(2) Paradigm uniformity effects:

- e.g. Scot. Eng. *tide/tied, brood/brewed*

(3) Dialect mixture:

- e.g. NE Am. Eng. *bad, mad, glad, sad*

Leaky phonology in spoken language?

- Phonemes are not always meaningless
- Some words contain sounds that are not part of a language's normal phoneme inventory
- Some allophonic rules have exceptions

Rationalisations of leaky phonology

(Phonemes are not always meaningless)

(1) Easy to justify treating onomatopoeia as marginal.

(2) Easy to justify treating sound symbolism as marginal, ***especially for linguists who are speakers of European languages.***

Rationalisations of leaky phonology

(Some words contain sounds that are not part of a language's normal phoneme inventory)

- Easy to justify treating interjections as marginal – ***but treatment is typically inconsistent.*** (Descriptions emphasise arbitrariness and language-specificity of interjections, yet they also ignore non-phonemic sounds because interjections are "expressive".)

Rationalisations of leaky phonology

(Some words contain sounds that are not part of a language's normal phoneme inventory)

- Similarly easy to ignore truly <u>foreign sounds</u>, but difficult to decide when a foreign sound has been integrated into borrowing language's phonology.

Rationalisations of leaky phonology

(Some words contain sounds that are not part of a language's normal phoneme inventory)

- More difficult to ignore unique phoneme in <u>"special words"</u> like *Allah* (but I know of no attempt to deal with the implications of such cases).

Rationalisations of leaky phonology

(Some allophonic rules have exceptions)

- Easy to treat allophonic exceptions involving <u>foreign words</u> as historical change "in progress". But again we have the problem of deciding when a foreign sound has been integrated into the borrowing language's phonology.

Rationalisations of leaky phonology

(Some allophonic rules have exceptions)

Japanese orthographic convention for dealing with exceptional allophony:

- e.g. Jap. /t/ before /i/

  tisshu ティッシュ　‘tissue (paper)’

  charitii チャリティー　‘charity’

Rationalisations of leaky phonology

(Some allophonic rules have exceptions)

- <u>Paradigm uniformity effects</u> are harder to ignore, but they can often be analysed in terms of allophony conditioned by morpheme boundaries or similar structural abstractions.

Rationalisations of leaky phonology

(Some allophonic rules have exceptions)

- “Dialect mixture” traditionally counts as a reason in its own right for ignoring the problems of exceptions to allophony. But there is a fundamental problem with falsifiability if we accept this; and in any case there is again the problem of deciding when a borrowed feature has been integrated.

Rationalisations of leaky phonology

(Some allophonic rules have exceptions)

- A recent attempt by Kiparsky (2014) to address this problem involves positing **intermediate phonological status** for some phenomena.
- He draws a distinction between "distinctiveness" (native speaker phonetic awareness) and "contrastiveness" (lexical phonological difference).

Rationalisations of leaky phonology

(Some allophonic rules have exceptions)

- Kiparsky 2014:

	contrastive	*non-contrastive*
distinctive	**phoneme**	**quasi-phoneme**
non-distinctive	**near-merger, incompl. neutr.**	**allophone**

Rationalisations of leaky phonology

(Some allophonic rules have exceptions)

- Some cases are difficult to classify even in Kiparsky's extended scheme, e.g. quasi-contrastive higher and lower mid vowels in Italian and French (Renwick and Ladd, work in progress).

Rationalisations of leaky phonology

(Some allophonic rules have exceptions)

- Kiparsky 2014: "The conjecture is that all phonemes arise as quasi-phonemes, and that all mergers pass through a near-merger stage".

This seems to imply that these categories only arise as transitional phenomena.

If phonology is “emergent”…

- Accept that there can be lexical exceptions to phonological generalisations, just like lexical exceptions to grammatical generalisations.
- Phonology is systematic internal structure for words or morphemes, as grammar is systematic internal structure for phrases and utterances.

If phonology is “emergent”…

- This view of phonology (which seems clearly justified for sign language) does not preclude iconicity, non-morphemic “meaning”, unique phonological features of specific words, etc.
- It does not require abrupt historical reanalysis of borrowings, or assume that intermediate status is only transitional.

If phonology is "emergent"…

- This view of phonology ***does*** put duality of patterning in a different light: Duality of patterning involves an intrinsically ***hierarchical*** relation, not two independent parallel structures, grammatical and phonological.
- "In the beginning was the word."

A broader view of "phonology"?

- This view is consistent with psycholinguistic work on speech perception and word recognition, and may make it possible to reconcile word-based exemplar models with the extensive evidence for the psychological reality of the phoneme.

A broader view of “phonology”?

- This view also points toward a theory of “phonology” independent of spoken medium:

“Phonology, construed broadly as an abstract theory of linguistic form, applies not only to speech but to other forms of communication (handwritten, printed, signed, etc.) as well.” (Kornai 2008)

Some references

Aronoff, Mark (2008). In the Beginning was the Word. *Language* 83: 803-830.

Brentari, Diane (1998). *A Prosodic Model of Sign Language Phonology.* Cambridge MA: MIT Press.

Fitch, W. T. and Jarvis, Erich (2013). Birdsong and other animal models for human speech, song, and vocal learning. In M. Arbib (ed.) *Language, music, and the brain: A mysterious relationship*. Strüngmann Forum Reports, vol. 10. Cambridge, MA: MIT Press.

Frishberg, Nancy (1975). Arbitrariness and iconicity: Historical change in American Sign Language. *Language* 51: 696-719.

Hjelmslev, Louis (1975). *Résumé of a theory of language* (translated and edited by Francis J. Whitfield). Madison: University of Wisconsin Press.

Hockett, Charles F. (1960). The origin of speech. *Scientific American* 203: 88-111.

Kaisse, Ellen (1985). *Connected Speech*. Academic Press.

Kiparsky, Paul (2014). New perspectives in historical linguistics. In C. Bowern and B. Evans (eds.) *The Routledge Handbook of Historical Linguistics.* Routledge.

Kornai, András (2008). *Mathematical linguistics*. Berlin: Springer.

Martinet, André (1949). La double articulation linguistique. *Travaux du Cercle Linguistique de Copenhague* 5: 30-37.

Nespor, Marina and Vogel, Irene (1986). *Prosodic Phonology.* Dordrecht: Foris Publications.

Sandler, Wendy; Aronoff, Mark; Meir, Irit; and Padden, Carol (2011). The gradual emergence of phonological form in a new language. *Natural Language and Linguistic Theory* 29: 503-543.

Sapir, Edward (1921). *Language.* New York: Harcourt Brace.

Selkirk, Elisabeth (1984). *Phonology and syntax: the relation between sound and structure.* Cambridge, MA: MIT Press.

Stokoe, William (1960). Sign language structure: An outline of the visual communication systems of the American Deaf. *Studies in Linguistics, Occasional Papers 8*, University of Buffalo. Reprinted 2005 in *Journal of Deaf Studies and Deaf Education* 10: 3-37.

Yip, Moira (2006). The search for phonology in other species. *Trends in Cognitive Sciences* 10: 442-446.

This talk is based on chapter 5 of my book *Simultaneous Structure in Phonology* (O.U.P., 2014).

DADDY, EDDY, NINNY, NANNY and BALDEY: Big Data for speech perception

Anne Cutler

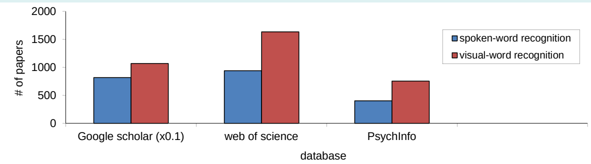
Radboud University Nijmegen

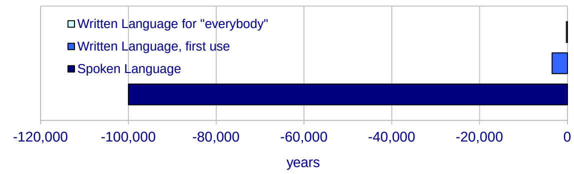
the MARCS institute

University of Western Sydney

MAX PLANCK INSTITUTE FOR PSYCHOLINGUISTICS

Research on spoken vs. visual language





Five data sets for speech perception research

Five data sets for speech perception research

- Speech perception research's scientific tradition: hypothesis-driven and experiment-based
- Big data of any kind notoriously hard to fund
- Often compiled by industry, or fully-funded government institutions
- Corpora: real life, undirected; but privacy issues.
- Who makes designed large data sets for speech perception research?

Five data sets for speech perception research

1. DADDY

Smits, R., Warner, N.L., McQueen, J.M. & Cutler, A. (2003). Unfolding of phonetic information over time: A database of Dutch diphone perception. *JASA*, **113**, 563-574.

http://www.mpi.nl/world/dcsp/diphones/index.html

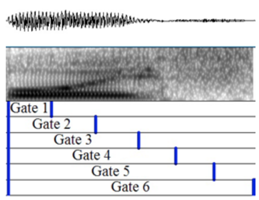
(Sound files [both full and gated], plus all responses from 18 listeners)

Why and how we collected this data set

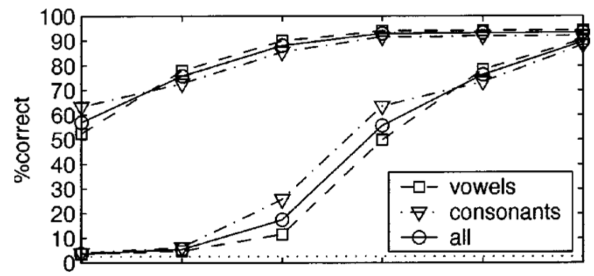
Our Aim: Data to support a more realistic front end for a spoken-word recognition model, for all phonemes of a language, in all contexts where they could possibly occur.

Experiment

- 2294 diphones: all possible within- or cross-word sequences of two Dutch phonemes including some stress variation (spoken by a single speaker)
- Each diphone gated to (mostly) 6 fragments (ending in square wave); Total = 13570 stimuli, randomised
- 18 listeners (judged phoneme 1 & 2)
- Total N responses per listener: 27140
- Average listener participation: 26 hrs
- Total database: 488520 data points

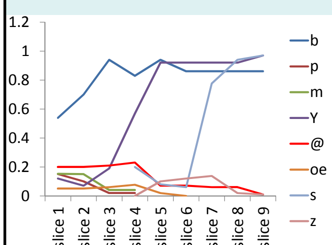


Orderly data!



% correct identifications for the diphones' Segment 1 (above), and segment 2 (below), across the 6 gated fragments of increasing size

DADDY data as front end for Shortlist B



The input *bus* presented to the model in 9 time slices; phoneme probabilities at each slice are computed from the DADDY data. These likelihoods continuously adjust the parameter P(E|W) which in turn adjusts P(W|E) i.e. the shortlist of lexical candidates.

likelihood prior p

$$P(Word_i \mid Evidence) = \frac{P(Evidence \mid Word_i) \times P(Word_i)}{\sum_{j=1}^{j=n} P(Evidence \mid Word_j) \times P(Word_j)}$$

posterior p normaliser

Five data sets for speech perception research

2. EDDY

Warner, N.L., McQueen, J.M. & Cutler, A. (2014). Tracking perception of the sounds of English. *JASA*, **135**, 2995-3006.

http://www.u.arizona.edu/~nwarner/WarnerMcQueenCutler.html

(Sound files and data files, for 20 listeners, as for DADDY)

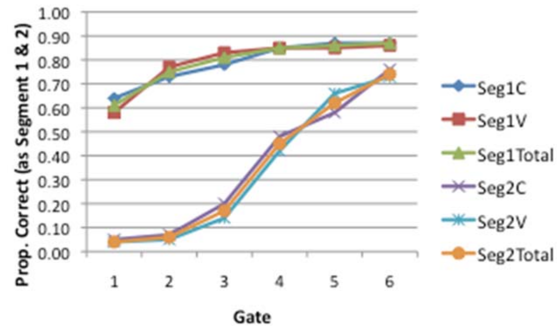
Why and how we collected this data set

Our Aim: Shortlist B works beautifully. An English front end would enable simulation of experiments in English, too.

Experiment

- All 2288 possible diphones of a variety of American English (spoken by a single speaker)
- Each diphone token again gated to (usually) 6 fragments (each ending in a square wave); Total: 13,464 stimuli
- 20 listeners judged all stimuli (1st and 2nd phoneme)
- Total number of responses per listener: 26928
- Average participation per listener: 33 one-hour sessions
- Total database: 538560 data points

More orderly data!



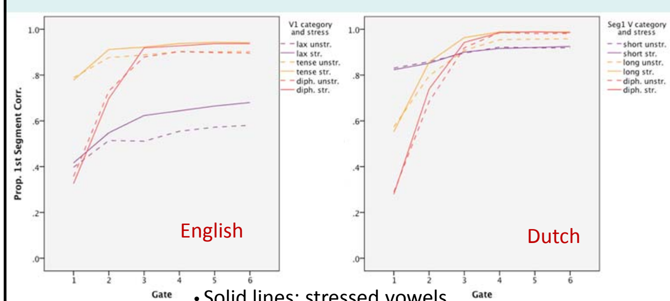
% correct identifications for the diphones' Segment 1 (above), and segment 2 (below), across the 6 gated fragments of increasing size

DADDY and EDDY can be compared, too

- Similar data sets, so: cross-language comparisons
- An example: stressed vs. unstressed vowels
- In Dutch, listeners attend to suprasegmental stress cues in recognising spoken words (e.g. *do-* from *DOminee* suffices to reject *domiNANT*)
- The same cues distinguish stressed from unstressed vowels in English, but English listeners rarely use them because inter-word distinctions rarely depend on it. (NB Dutch listeners to English do use the English cues!!)
- Are stress effects on vowel identification similar in the two languages?

(Cooper, Cutler & Wales, *Lg&Sp* 2002; Donselaar, Koster & Cutler, *QJEP* 2005; Cutler, *JASA* 2009)

Vowel identification in English & Dutch



- Solid lines: stressed vowels.
- Dashed lines: unstressed vowels.

Dutch: little stress effect (They are used to differently stressed vowels)
English: big effect (They don't expect vowels in multiple stress versions)

Five data sets for speech perception research

3. NINNY

Cutler, A., Weber, A., Smits, R. & Cooper, N. (2004). Patterns of English phoneme confusions by native and non-native listeners. *JASA*, **116**, 3668-3678.

http://www.mpi.nl/people/cutler-anne/research

(Full identification response set from 16 native [American English] and 16 non-native [Dutch] listeners given American English CV or VC input)

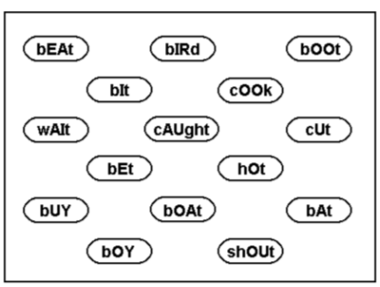
Why and how we collected this data set

Our Aim: Why exactly is non-native listening in noise so hard? If all predictability (lexical, any kind of contextual) is removed, do non-native listeners still suffer more from noise interference than native listeners? i.e. Do they always need better low-level evidence; or are they just less able to profit from higher-level predictability to recover from interference?

Experiment

- All possible CV and VC sequences of AmEng; 645 items
- In 3 levels of multi-talker babble noise (0, 8, 16 dB SNR)
- 32 listeners (16 each AmEng, Dutch) identified each phoneme of each syllable separately (3870 trials each)
- Total data set: 123840 data points

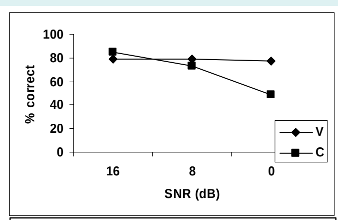
Response display



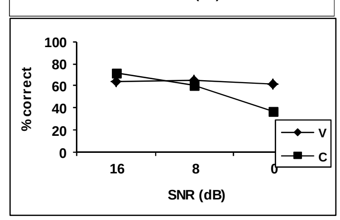
Separate displays for vowels, initial consonants and final consonants

Results

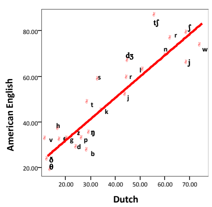
Native listeners (American English):



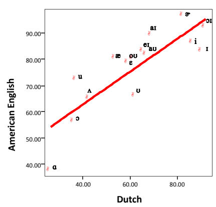
Non-native listeners (Dutch):



Results: Vowel and consonant identification



consonants

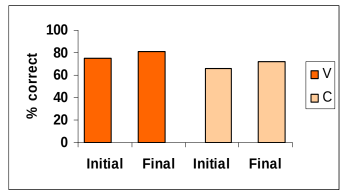


vowels

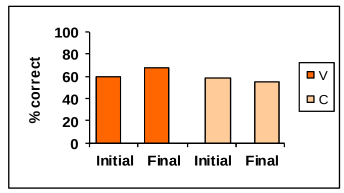
Highly significant positive correlation (r = .91) between percent correct recognition per phoneme by native (vertical axis) and non-native listeners (horizontal axis)

Results: Position effects

Native listeners (American English):



Non-native listeners (Dutch):



Why is L2 listening in noise so hard?

- Noise masks non-native listening and native listening similarly
- The extra difficulty of non-native listening in noise is not due to phoneme identification problems alone
- It's because non-native listeners can't recover from these problems

Five data sets for speech perception research

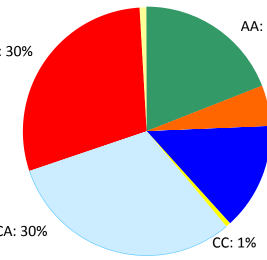
4. NANNY

Johnson, E.K., Lahey, M., Ernestus, M. & Cutler, A. (2013). A multimodal corpus of speech to infant and adult listeners. *JASA*, **134**, EL534-540.

Previously: Language input from 6 to 9 months

Van de Weijer (1999) *"Language Input for Word Discovery"*
3 months, all input heard by a single infant
3 weeks (85000 utterances) fully analysed

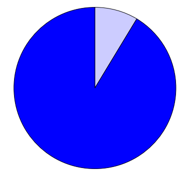
Previously: Language input from 6 to 9 months



What were the 85000 utterances?

1. Total Speaking Time: 2.56 hours per day
2. Speech types, by A(dult), C(hild) to A, C, I(nfant):

= 21.5 min.



isolated words (8.67%)
words in longer utterances

(Van de Weijer, 1999)

➡ Infants hear mostly continuous speech

Why and how we collected this data set

Our Aim: Answer some questions raised by existing corpora and provide relevant evidence on early word form acquisition.

Data Set

- 65 play sessions (33 hours of speech interaction) involving 28 triads, each of an 11-month-old infant with 2 caregivers
- Audio and (double) video record
- In part of the sessions, caregivers attempted to teach their infant new words
- In other parts, the caregivers interact with an experimenter and/or with each other or the infant

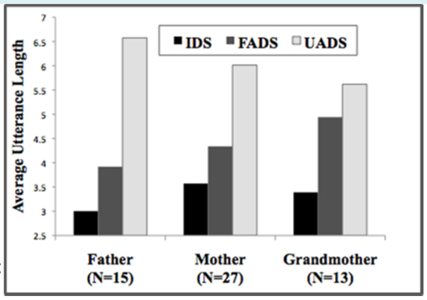
A word teaching example

The words were: a noun (e.g. *cactus*), a proper name (e.g. *Tigo*), a verb (e.g. *buigen* 'bow') and an adjective (e.g. *glanzend* 'shiny').

Double-view video allows eye gaze to be determined.

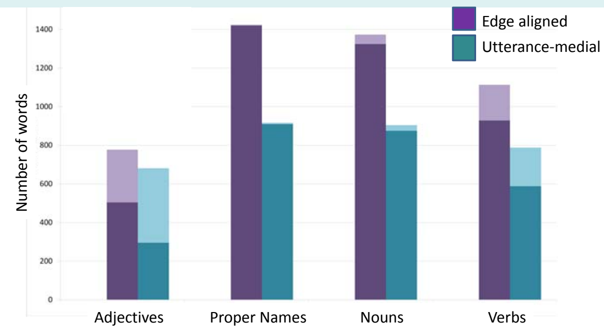
"Corpus" analysis: IDS vs. ADS

speaker:



Consistent with cross-corpus asymmetries, within this one corpus the difference between IDS and F(amiliar)ADS is much smaller than that between IDS and U(nfamiliar)ADS.

"Experimental" analysis: Word form segmentation



In agreement with the Edge Hypothesis, caregivers positioned target words at utterance edges. (Johnson, E.K., Seidl, A., Tyler, M.D. [2014]. The edge factor in early word segmentation: utterance-level prosody enables word form extraction by 6-month-olds. *PLoS ONE*, 9, e83546.)

Five data sets for speech perception research

5. BALDEY

Ernestus, M. & Cutler, A. BALDEY: A database of auditory lexical decisions. *Quarterly Journal of Experimental Psychology*, revision submitted, 2014.

http://www.mirjamernestus.nl/Ernestus/Baldey/index.html

(Sound files and Praat scripts for all 5541 items, and the full data set [accuracy, RTs] from 20 listeners)

Why and how we collected this data set

Our Aim: Data to support modelling of the lexical decision task and of recognition of spoken words of varying structure. Well-understood task, but little data across types of words.

Experiment

- 5541 items; 2780 real Dutch words, 2761 pseudo-words
- 20 participants (10 M 10 F). 10 5-part sessions each.
- Realistic variation in word class (verb [regular, irregular], noun, adjective), length (1 to 5 syllables), morphology (stem+deriv 27.7%, stem+infl. 21.9%, stem+2 affixes 13.3%, simple 18.4%, compound 13.5%, compound+affix 5.2%)
- Pseudo-words (a) matched to real words on structural factors; (b) phonologically plausible
- 110420 timed responses

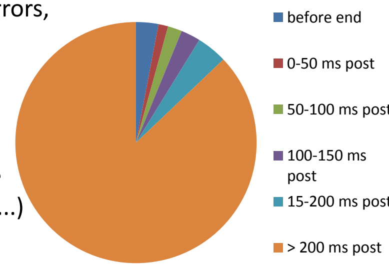
Lexical decision

The nature of the lexical decision task

1. Words are heard in isolation. (So: no contextual support)
2. There are both words and non-words.

Thus to avoid making errors, listeners must be sure they have heard each entire stimulus item.

(even a beginning like *televisio-* might become a nonword with *-d* or *-z*...)

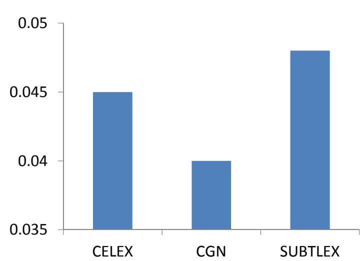


Our data show that our listeners performed the task appropriately.

Comparing corpora via this data set!

Data set offers many analysis options.
We include frequency measures from several corpora: CELEX, Corpus of Spoken Dutch (CGN), SUBTLEX.

Averaging across all word types, correlation of log RT measured from word offset with log word-form frequency in each of these corpora:



Five data sets for speech perception research

- Speech perception research's scientific tradition: hypothesis-driven and experiment-based
- Big experimental data sets allow testing of many hypotheses beyond those that motivated them
- Over to you....

LabPhon 14 - The 14th Conference on Laboratory Phonology

Tokyo, July 25-27, 2014

Pre-Conference Colloquium

Date

- **Thursday July 24, 2014, 14:30-17:30**
- **Registration for the main conference will also be possible on this day from 13:30-17:30.**

Guest Speakers

- **Bob Ladd (University of Edinburgh)**

"Leaky phonology and the design of language" [slides]

Abstract:
Recent work on sign language - notably the work of Sandler and colleagues on Al-Sayyid Bedouin Sign Language - has shown that phonology emerges in a new language. There does not appear to be a clear transition point from not having phonology to having it. Rather, the phonological organisation of signs becomes gradually more systematic, and even mature sign languages continue to show a variety of features that are difficult to analyse in terms of a smallish finite inventory of phonological elements. In this paper I suggest that the same is true of spoken language phonology as well, and that spoken languages exhibit phonological phenomena analogous to the hard-to-analyse features in sign phonologies. Sapir famously said that "all grammars leak". We are used to thinking that phonology does not leak, and that all phonetic aspects of an utterance can be analysed in terms of a language's finite inventory of phonological elements, independently of grammatical structure. "Leaky" aspects of spoken language phonology are usually idealised out of consideration in various ways (e.g. as paralinguistic or expressive, as dialect mixture or unassimilated borrowing, as historical change in progress), or simply ignored. I discuss several such phenomena, and suggest that they are comparable to what we find in sign languages. The existence of such phenomena argues against a conception of language design in which phonology is autonomous and unconnected to grammar and meaning (the design feature called "duality of patterning" by Hockett and "double articulation" by Martinet). Rather, phonology is in the first instance a property of the internal structure of signs (or words or morphemes), and it is not always possible or desirable to distinguish phonological from grammatical aspects of this internal structure. If phonology is viewed in this light, it is not surprising that it can "leak" like grammar.

- **Anne Cutler (University of Western Sydney)**

"DADDY, EDDY, NINNY, NANNY and BALDEY: Big Data for speech perception" [slides]

Abstract:
In its concluding decade, the Comprehension Group at the Max Planck Institute for Psycholinguistics created several very large sets of speech perception data, all of which are publicly available for the use of any interested parties (such as, for instance, the laboratory phonology community). The Database of Dutch Diphones (DADDY; Smits, Warner, McQueen & Cutler, JASA, 2003) comprises more than a half a million identification responses, by 18 listeners presented with gated fragments of diphones representing every phonetic segment of Dutch in every possible left and right phonetic

context. Even larger is DADDY's English counterpart, the English Diphones Database (EDDY; Warner, McQueen & Cutler, JASA, 2014) which used the same technique with American English input and 22 listeners. The Noise-masked Identifications by Native and Non-Native listeners data set (NINNY; Cutler, Weber, Smits & Cooper, JASA, 2004) contains identification responses by American English (native) and Dutch (non-native) listeners to American English vowels and consonants presented under three levels of noise masking. The other two data sets do not concern phonemic identifications. One is a large corpus of input to 11-month-old infants learning Dutch, with speech from multiple caregivers to the infant and to other adults (NANNY; Johnson, Lahey, Ernestus & Cutler, JASA, 2013). Finally, the BALDEY data set (Ernestus & Cutler, submitted 2014) contains lexical decisions to 2780 spoken words of Dutch and a like number of spoken nonwords. Indicative analyses will illustrate the possibilities of these multifaceted data sets.

Afterwards

17:30- : Informal wine reception

Back to home Back Top

Sitemap

LabPhon 14 - The 14th Conference on Laboratory Phonology

Tokyo, July 25-27, 2014

Satellite Workshop

- **Date:**
  Monday July 28, 2014.
- **Theme:**
  Gestural coordination within and between speakers in first language phonological acquisition.
- **Organizers:**
  Felicitas Kleber (Ludwig-Maximilian-Universität München)
  Mary E. Beckman (Ohio State University).
- **For more information, please visit the official workshop website.**

Satellite Events
- Pre-Conference Colloquium
- Satellite Workshop
- Call for Proposals

Back to home Back Top

Sitemap

LabPhon 14 - The 14th Conference on Laboratory Phonology

Tokyo, July 25-27, 2014

Call for Proposals

The Call for Satellite Workshop Proposals has been closed.

Call for Satellite Workshop Proposals

The conference venue can accommodate up to 4 satellite workshops. The days available for workshops are Thursday, 24 July, and Monday, 28 July. To propose a satellite workshop, please prepare a brief description (no more than 2 pages) including the following information:

- the workshop title
- the name(s) and affiliation(s) of the organizer(s)
- a 1-paragraph description of the workshop's purpose
- the date you prefer
- the total amount of time required, from beginning to end
- the expected number of presentations (both oral and poster)
- an estimate of the number of participants

Important dates:

- Proposal submission period: December 1, 2013 - January 31, 2014
- Notification of acceptance: March 15, 2014

Submission of Proposals

Submit the proposal (either as a Word file or as a PDF file) by attaching it to an e-mail message addressed to:

labphon14@ninjal.ac.jp

Please note: If the workshop involves collecting a fee and/or registration information from participants, the workshop organizer(s) must make the necessary arrangements, separate from the registration and payment process for the main LabPhon14 conference.

If your workshop proposal is accepted, please understand that NINJAL can provide only the meeting space. All other workshop-related tasks (planning, budget, publicity, etc.) are the responsibility of the organizer(s).

Visa Information for Participants from Abroad

The regulations for entering Japan as a visitor differ significantly depending on the country that issued your passport. For many countries, no visa is required to enter Japan as a tourist and stay up to 90 days. However, for many other countries, a visa is required even for a short stay. If you are not sure about whether you need a visa, please contact your nearest Japanese embassy or consulate.

Satellite Events
- Pre-Conference Colloquium
- Satellite Workshop
- Call for Proposals

10-2 Midori-cho, Tachikawa City, Tokyo, Japan 190-8561

This superficial sketch of Japanese orthography and phonology is an excerpt from:

Vance, Timothy J. 2001. Japanese. *Facts about the World's Languages*, ed. by Jane Garry and Carl Rubino, 345–349. New York & Dublin: H. W. Wilson.

Orthography and Basic Phonology

The Japanese writing system is generally regarded as the modern world's most complex. Roughly speaking, native and Sino-Japanese nouns are written with Chinese characters (*kanji*), as are the invariant portions of most inflected words (verbs and adjectives). Inflectional endings, postpositions, and many adverbs are spelled out in a cursive syllabary (*hiragana*), and non-Chinese loanwords are spelled out in a squareish syllabary (*katakana*). There are, however, many exceptions to these broad generalizations. A single Chinese character often has two or more possible readings, the intended reading being determined by the context. The *hiragana* and *katakana* originated as abbreviated versions of certain Chinese characters that had come to be used as phonograms.

There are two principal systems of Romanization, and a variant of the more popular Hepburn system is used here, with long vowels indicated by double letters (except in place-names mentioned in the text, where macrons are used instead). The chart on the next page shows the individual symbols of *hiragana* and *katakana* in the traditional right-to-left arrangement of top-to-bottom lines. The pronunciation of each symbol is given in Romanization. The many possible Japanese syllables not included in the chart are written by using diacritics and/or by combining two or three symbols.

All consonants listed in the Table l occur initially in CV syllables, and many can also occur initially in C*y*V syllables. A syllable-final consonant must be one of the two moraic consonants, which count as separate beats in poetic meter. The moraic nasal is always Romanized as *n*, but its pronunciation assimilates to the immediately following sound and varies widely, such as a bilabial [m] before bilabials, as in *onpa* 'soundwave', a nasalized glide [ɰ̃] before *s*, as in *kansa* 'inspection'; before a pause, it is usually uvular [ɴ]. Before a vowel or *y*, a moraic nasal is followed by an apostrophe in Romanization to distinguish it from a syllable-initial alveolar nasal: *an'i* 'easy', *ani* 'older brother'. The moraic obstruent combines with an immediately following obstruent to form a phonetically long obstruent and is Romanized as a copy of that following obstruent, as in *rappa* 'bugle' and *dassen* 'derailment'. Except in recent loanwords, the moraic obstruent occurs only before voiceless obstruents.

The voiceless stops and affricates are unaspirated. Some speakers can substitute a velar nasal [ŋ] for a word-medial voiced velar stop *g* in a large subset of the vocabulary. The alveolar affricates are *ts* [ts] and *z* [dz], but *z* is pronounced as a fricative [z] word-medially unless preceded by the moraic nasal. The alveo-palatal affricates are *ch* [cɕ] and *j* [ɟʑ]. The fricative *f* [ɸ] was until recently an allophone of *h*, occurring only before *u*, but *f* now appears before all vowels in loanwords, for example, *ferii* 'ferry'. The fricative *h* is pronounced as palatal [ç] before *y* or *i*. The alveo-palatal fricative is *sh* [ɕ]. The liquid is a tap *r* [ɾ], and the velar glide *w* is often described as unrounded [ɰ], although it shows lip compression (not protrusion) in careful pronunciation. Even in recent loanwords, *t* and *d* do not occur before *u*, *s* and *z* do not occur before *y* or *i*, *h* does not occur before *u*, *y* does not occur before *i*, and *w* does not occur before *u*.

The high back vowel *u* is often described as unrounded [ɯ], but like the glide *w* it shows lip compression in careful pronunciation. All five vowels occur both short and long, and a syllable with a long vowel counts as two moras. All two-vowel sequences are possible, although some are quite rare; if the second vowel is *e*, *o*, or *a*, it is in a separate syllable from the first, but if the second vowel is *i* or *u*, it may be the second mora of a long syllable. Short *i* and *u* are frequently devoiced between consonants or between a voiceless consonant and a pause.

Japanese has a pitch-accent system, and the location of accent, manifested as a drop from high to low pitch, is the contrastive feature of the system. The presence or absence of an accent and its location in a word are essential to determining the intonation contour of a phrase containing that word. The possible locations for accent are least constrained in nouns. A noun may be accented on any syllable, or it may be unaccented, as in *ha'shi* 'chopsticks', *hashi'* 'bridge' and *hashi* (unaccented) 'edge'. Accent is not marked in Japanese orthography or in Romanization.

Hiragana

ん [moraic]n	わ wa	ら ra	や ya	ま ma	は ha	な na	た ta	さ sa	か ka	あ a
		り ri		み mi	ひ hi	に ni	ち chi	し shi	き ki	い i
		る ru	ゆ yu	む mu	ふ fu	ぬ nu	つ tsu	す su	く ku	う u
		れ re		め me	へ he	ね ne	て te	せ se	け ke	え e
	を (w)o	ろ ro	よ yo	も mo	ほ ho	の no	と to	そ so	こ ko	お o

Katakana

ン [moraic]n	ワ wa	ラ ra	ヤ ya	マ ma	ハ ha	ナ na	タ ta	サ sa	カ ka	ア a
		リ ri		ミ mi	ヒ hi	ニ ni	チ chi	シ shi	キ ki	イ i
		ル ru	ユ yu	ム mu	フ fu	ヌ nu	ツ tsu	ス su	ク ku	ウ u
		レ re		メ me	ヘ he	ネ ne	テ te	セ se	ケ ke	エ e
	ヲ (w)o	ロ ro	ヨ yo	モ mo	ホ ho	ノ no	ト to	ソ so	コ ko	オ o

Table l: Consonants

	Bilabial	Alveolar	Alveo-palatal	Velar	Glottal
Stops Voiceless Voiced	 p b	 t d		 k g	
Fricatives	f	s	sh		h
Affricates Voiceless Voiced		 ts z	 ch j		
Nasals	m				
Liquid		r			
Glides			y	w	

Table 2: Vowels

	Front	Back
High	i	u
Mid	e	o
Low		a

Thursday, July 24	Friday, July 25	Saturday, July 26	Sunday, July 27	Monday, July 28
	8:30-9:20 Registration			Satellite Workshop: MORNING SESSION
	9:20-9:30 Welcome			
	9:30-10:50 ORAL SESSION 1	9:10-10:30 ORAL SESSION 5	9:10-10:30 ORAL SESSION 9	
	Coffee Break	Coffee Break	Coffee Break	
	11:10-12:10 ORAL SESSION 2	10:50-12:10 ORAL SESSION 6	10:50-12:10 ORAL SESSION 10	
	Lunch	Lunch (LP Journal meeting)	Lunch	Lunch
13:30-17:30 LABPHON 14 Registration	13:10-14:30 ORAL SESSION 3	13:10-14:30 ORAL SESSION 7	13:10-15:00 POSTER SESSION 3 (with coffee)	
14:30-17:30 COLLOQUIUM - D. Robert Ladd - Anne Cutler	14:30-16:20 POSTER SESSION 1 (with coffee)	14:30-16:20 POSTER SESSION 2 (with coffee)	15:10-16:10 ORAL SESSION 11	Satellite Workshop: AFTERNOON SESSION
17:30~ Wine reception	16:30-18:10 ORAL SESSION 4	16:30-17:50 ORAL SESSION 8	16:10-16:30 GENERAL DISCUSSION	
			16:30-16:35 Farewell	
	18:10-18:40 ALP General Meeting	18:00-20:00 Banquet and Fireworks		~ 17:30

LabPhon 14

The 14th Conference on
Laboratory Phonology

July 25-27, NINJAL, Tokyo

Table of Contents

Message from the LabPhon 14 Organizers

Welcome to LabPhon14!

We are delighted to be able to host the 14th Conference on Laboratory Phonology at NINJAL (the National Institute for Japanese Language and Linguistics), and we would like to welcome all the participants to Tachikawa.

The response to the call for papers was overwhelming, and we are grateful to the many scholars who gave so generously of their time to review abstracts. The number of people wanting to participate in the meeting also surpassed all expectations, forcing us to close registration much earlier than anticipated so as not to exceed the capacity of the venue.

As you probably know, this is not the best time of year to visit Japan. Like most of the country, Tokyo in late July is almost guaranteed to be hot and extremely humid. There are occasional thunderstorms, but they typically just raise the humidity even higher without appreciably reducing the temperature. Please take care to avoid heat stroke and stay well hydrated.

If you have any questions or concerns, feel free to ask any of the organizers or conference staff members. If the person you ask does not have the information you need, he or she will find someone who does. We will do our best to make your stay as enjoyable and productive as possible.

Haruo Kubozono, Kikuo Maekawa, Tim Vance
LabPhon14 Conference Organizers

NINJAL
National Institute for Japanese Language and Linguistics

Organization

Executive Council of the Association for Laboratory Phonology (term of office)

Bob Ladd, President (2012-2014)
Martine Grice, Vice-President (2012-2014)
Jennifer Cole, Editor-in-Chief of Laboratory Phonology (2010-2015)
Matt Goldrick, Secretary (2012-2016)
Caroline Smith, Treasurer (2012-2014)
Grzegorz Dogil, Organizing committee LabPhon 13 (2010-2014)
Haruo Kubozono, Organizing committee LabPhon 14 (2012-2016)
Mary E. Beckman, Councilor-at-large (2012-2014)
Ioana Chitoran, Université Paris 7, Councilor-at-large (2012-2016)
Taehong Cho, Councilor-at-large (2012-2016)
José Ignacio Hualde, Councilor-at-large (2012-2014)

Local Organizing Committee

Haruo Kubozono (Dept. of Linguistic Theory and Structure, NINJAL)
Kikuo Maekawa (Dept. of Corpus Studies, NINJAL)
Timothy Vance (Dept. of Linguistic Theory and Structure, NINJAL)
Mafuyu Kitahara (Waseda University)
Keiichi Tajima (Hosei Univerisity)
Kiyoko Yoneyama (Daito Bunka University)
Mikio Giriko (JSPS/NINJAL)
Hyun Kyoung Hwang (Dept. of Linguistic Theory and Structure, NINJAL)
Clemens Poppe (The University of Tokyo/Dept. of Linguistic Theory and Structure, NINJAL)
Mee Sonu (Dept. of Linguistic Theory and Structure, NINJAL)
Izumi Takiguchi (Sophia University/Dept. of Linguistic Theory and Structure, NINJAL)
Junko Yoneda (Dept. of Linguistic Theory and Structure, NINJAL)

Cooperating Organizations

The Acoustical Society of Japan
The Linguistic Society of Japan
The Phonetic Society of Japan
The Phonological Society of Japan

Sponsor

National Institute of Japanese Language and Linguistics (NINJAL)

Invited Speakers

Yasuharu Den (Chiba University/NINJAL)
Topic: Some cognitive factors behind vowel lengthening in spontaneous Japanese: A corpus-based study
Commentator: Shu-Chuan Tseng (Academia Sinica)

Carlos Gussenhoven (Radboud University Nijmegen)
Topic: On establishing the existence of word stress
Commentator: Aditi Lahiri (University of Oxford)

Mark Hasegawa-Johnson (University of Illinois at Urbana-Champaign)
Topic: Labeling in the wild: Crowdsourcing versus categorical perception
Commentator: Natasha Warner (University of Arizona)

Reiko Mazuka (Duke University/RIKEN Brain Science Institute)
Topic: Infant-directed speech as a window into the dynamic nature of phonology
Commentator: Catherine Best (University of Western Syndey)

Satellite Events

- ***Pre-Conference Colloquium*** (Thursday July 24, 2014, 14:30-17:30; NINJAL)
  - **D. Robert Ladd (University of Edinburgh)**

    Topic: Leaky phonology and the design of language

    Abstract: Recent work on sign language - notably the work of Sandler and colleagues on Al-Sayyid Bedouin Sign Language - has shown that phonology emerges in a new language. There does not appear to be a clear transition point from not having phonology to having it. Rather, the phonological organisation of signs becomes gradually more systematic, and even mature sign languages continue to show a variety of features that are difficult to analyse in terms of a smallish finite inventory of phonological elements. In this paper I suggest that the same is true of spoken language phonology as well, and that spoken languages exhibit phonological phenomena analogous to the hard-to-analyse features in sign phonologies.

    Sapir famously said that "all grammars leak". We are used to thinking that phonology does not leak, and that all phonetic aspects of an utterance can be analysed in terms of a language's finite inventory of phonological elements, independently of grammatical structure. "Leaky" aspects of spoken language phonology are usually idealised out of consideration in various ways (e.g. as paralinguistic or expressive, as dialect mixture or unassimilated borrowing, as historical change in progress), or simply ignored. I discuss several such phenomena, and suggest that they are comparable to what we find in sign languages. The existence of such phenomena argues against a conception of language design in which phonology is autonomous and unconnected to grammar and meaning (the design feature called "duality of patterning" by Hockett and "double articulation" by Martinet). Rather, phonology is in the first instance a property of the internal structure of signs (or words or morphemes), and it is not always possible or desirable to distinguish phonological from grammatical aspects of this internal structure. If phonology is viewed in this light, it is not surprising that it can "leak" like grammar.

- **Anne Cutler (University of Western Sydney)**
  Topic: DADDY, EDDY, NINNY, NANNY and BALDEY: Big Data for speech perception
  Abstract: In its concluding decade, the Comprehension Group at the Max Planck Institute for Psycholinguistics created several very large sets of speech perception data, all of which are publicly available for the use of any interested parties (such as, for instance, the laboratory phonology community). The Database of Dutch Diphones (DADDY; Smits, Warner, McQueen & Cutler, JASA, 2003) comprises more than a half a million identification responses, by 18 listeners presented with gated fragments of diphones representing every phonetic segment of Dutch in every possible left and right phonetic context. Even larger is DADDY's English counterpart, the English Diphones Database (EDDY; Warner, McQueen & Cutler, JASA, 2014) which used the same technique with American English input and 22 listeners. The Noise-masked Identifications by Native and Non-Native listeners data set (NINNY; Cutler, Weber, Smits & Cooper, JASA, 2004) contains identification responses by American English (native) and Dutch (non-native) listeners to American English vowels and consonants presented under three levels of noise masking. The other two data sets do not concern phonemic identifications. One is a large corpus of input to 11-month-old infants learning Dutch, with speech from multiple caregivers to the infant and to other adults (NANNY; Johnson, Lahey, Ernestus & Cutler, JASA, 2013). Finally, the BALDEY data set (Ernestus & Cutler, submitted 2014) contains lexical decisions to 2780 spoken words of Dutch and a like number of spoken nonwords. Indicative analyses will illustrate the possibilities of these multifaceted data sets.

■ ***Satellite Workshop*** (Monday July 28, 2014; NINJAL)

Theme: Gestural coordination within and between speakers in first language phonological acquisition.

Organizers: Felicitas Kleber (Ludwig-Maximilians-Universität München, Germany)
Mary E. Beckman (Ohio State University, USA)

This satellite workshop, to be held in conjunction with LabPhon 14, addresses the challenges of investigating the role of gestural coordination in first language phonological acquisition, with coordination construed broadly to encompass both the coordination of gestures within the individual infant or child and the coordination of gestures between the individual and others in social interactions such as the act of imitation. The workshop will focus on both types of coordination, in an afternoon session with submitted papers on the quantification of data obtained with experimental techniques following a morning session with invited speakers talking about imitation and learning from infancy through early childhood. The goal is to discuss novel experimental methods in phonological acquisition research and to relate models of imitation and learning in adults to models of imitation and learning in infants and toddlers. **URL:http://phonetik.uni-muenchen.de/labphon14-satellite/index.html**

List of Reviewers

We sincerely appreciate their contributions.

Eleanora Albano
Jalal-eddin Al-Tamimi
Diana Archangeli
Amalia Arvaniti
Harald Baayen
Mary Beckman
Tessa Bent
Pier Marco Bertinetto
Sonya Bird
Dani Byrd
Loana Chitoran
Taehong Cho
Cynthia Clopper
Abby Cohn
Jennifer Cole
Onno Crasborn
Mariapaola D'Imperio
Lisa Davidson
Barbara Davis
Kenneth de Jong
Didier Demolin
Katherine Demuth
Yasuharu Den
Laura Dilley
Gerry Docherty
Grzegorz Dogil
Ulrike Domahs
Frank Eisner
Mirjam Ernestus
John Esling
Carsten Eulitz
Zsuzsanna Fagyal
Caroline Féry
Colleen Fitzgerald
Cécile Fougeron
Paul Foulkes
Claudia Friedrich
Stefan Frisch
Susanne Fuchs
Adamantios Gafos
Cedric Gendrof
Fiona Gibbon
Barbara Fivela Gili
Matt Goldrick
Matt Gordon
Martine Grice
Carlos Gussenhoven
John Hajek
Silke Hamann
Mike Hammond
Sharon Hargus
Jonathan Harrington
Ingo Hertrich
Barry Heselwood
Julia Hirschberg
Philip Hoole
Jose Hualde
Elizabeth Hume
Khalil Iskarous
Ewa Jacewicz
Thomas Jacobsen
Matthias Jilka
John Kingston
Alexei Kochetov
Frank Kügler
Barbara Kühnert
Aditi Lahiri
Erika Levy
Ian Maddieson
Egidio Marsico
Gaurav Mathur
Reiko Mazuka
Joyce McDonough
Yohann Meynadier
Alexis Michaud
Jeff Mielke
Amanda Miller
Bernd Moebius
Christine Mooshammer
Doris Mücke
Benjamin Munson
Scott Myers
James Myers
Noel Nguyen
Richard Ogden
Jaye Padgett
Mary Pearce
François Pellegrino
Pascal Perrier
Melissa Redford
Henning Reetz
Rachid Ridouane
Joaquìn Romero
Yvan Rose
Rebecca Scarborough
Stefanie Shattuck-Hufnagel
Pat Shaw
Linda Shockey
Caroline Smith
Donca Steriade
Jane Stuart-Smith
Marc Swerts
Shu-Chuan Tseng
Alice Turk
Nathalie Vallèe
Ruben van de Vijver
Els van der Kooij
Shravan Vasishth
Eric Vatikiotis-Bateson
Mario Vayra
Petra Wagner
Michael Walsh
Natasha Warner
Janet Watson
Pauline Welby
Doug Whalen
Richard Wiese
Sherman Wilcox
Colin Wilson
Alan Yu
Chakir Zeroual
Jie Zhang
Marzena Zygis

Previous Meetings

1st	LabPhon1, Columbus, OH, USA	(1987)
2nd	LabPhon2, Edinburgh, UK	(1989)
3rd	LabPhon3, Los Angeles, CA, USA	(1991)
4th	LabPhon4, Oxford, UK	(1993)
5th	LabPhon5, Evanston, IL, USA	(1996)
6th	LabPhon6, York, UK	(1998)
7th	LabPhon7, Nijmegen, the Netherlands	(2000)
8th	LabPhon8, New Haven, CT, USA	(2002)
9th	LabPhon9, Urbana, IL, USA	(2004)
10th	LabPhon10, Paris, France	(2006)
11th	LabPhon11, Wellington, New Zealand	(2008)
12th	LabPhon12, Albuquerque, NM, USA	(2010)
13th	LabPhon13, Stuttgart, Germany	(2012)

Statistics

Abstracts

	Submitted	Accepted	%
Oral	262	27 (Student 6, Non-Student 21)	10.3
Poster		116 (Student 47, Non-Student 69)	44.3
Total	262	143	54.6

Registrations

	Student	Non-Student	Total
Member	11	47	58
Non-Member	63	113	176
Total	74	160	234

Authors and Participant

	Accepted Authors by Country (First authors only)	Participants
Australia	13	19
Belgium	1	1
Canada	9	19
France	8	7
Germany	13	16
Greece	1	0
Hong Kong SAR	3	4
Japan	9	74
Korea	1	5
New Zealand	1	1
Singapore	3	2
Spain	1	0
Sweden	1	1
Switzerland	1	1
Taiwan	6	9
Thailand	1	1
The Netherlands	9	11
United Kingdom	19	14
USA	43	49
Total	143	234

Conference Schedule

Thursday, July 24	Friday, July 25	Saturday, July 26	Sunday, July 27	Monday, July 28
	8:30-9:20 Registration			
	9:20-9:30 Welcome			
	9:30-10:50 ORAL SESSION 1	9:10-10:30 ORAL SESSION 5	9:10-10:30 ORAL SESSION 9	Satellite Workshop: MORNING SESSION
	Coffee Break	Coffee Break	Coffee Break	
	11:10-12:10 ORAL SESSION 2	10:50-12:10 ORAL SESSION 6	10:50-12:10 ORAL SESSION 10	
	Lunch	Lunch (LP Journal meeting)	Lunch	Lunch
13:30-17:30 -Registration	13:10-14:30 ORAL SESSION 3	13:10-14:30 ORAL SESSION 7	13:10-15:00 POSTER SESSION 3 (with coffee)	
14:30-17:30 COLLOQUIUM - D. Robert Ladd - Anne Cutler	14:30-16:20 POSTER SESSION 1 (with coffee)	14:30-16:20 POSTER SESSION 2 (with coffee)	15:10-16:10 ORAL SESSION 11	Satellite Workshop: AFTERNOON SESSION
	16:30-18:10 ORAL SESSION 4	16:30-17:50 ORAL SESSION 8	16:10-16:30 GENERAL DISCUSSION	
17:30~ -Wine reception			16:30-16:35 Farewell	
	18:10-18:40 ALP General Meeting	18:00-20:00 Banquet and Fireworks		

LabPhon14 PROGRAM (ORAL PRESENTATIONS)

THURSDAY, July 24

14:30 –17:30 Colloquium: D. Robert Ladd (University of Edinburgh)
Anne Cutler (University of Western Sydney)

13:30 –17:30 Registration / 17:30– wine reception

FRIDAY, July 25

08:30 – 09:20 **Registration**

09:20 – 09:30 **Welcome**

SESSION 1 Chair: Shigeto Kawahara (Keio University)

09:30 –10:20 **[Invited] Infant-directed speech as a window into the dynamic nature of phonology**
Reiko Mazuka (RIKEN Brain Science Institute)

10:20 –10:50 **Comments and Discussion**: Catherine T. Best (University of Western Sydney)

10:50 –11:10 **BREAK**

SESSION 2 Chair: Emiko Kaneko (University of Aizu)

11:10 –11:30 **Children's imitation of coarticulatory patterns in different prosodic contexts**
Felicitas Kleber (Ludwig-Maximilians-University), Sandra Peters (Ludwig-Maximilians-University)

11:30 –11:50 **An investigation into articulatory adaptation during acoustic mimicry of postvocalic /r/**
Eleanor Lawson (Queen Margaret University), James M. Scobbie (Queen Margaret University), Jane Stuart-Smith (University of Glasgow)

11:50 –12:10 **Dolls are prissy: Preschoolers' toy preferences predict medial /t/ production**
Jacqui Nokes (University of Canterbury)

12:10 –13:10 **LUNCH**

SESSION 3 Chair: Mariko Kondo (Waseda University)

13:10 –13:30 **Constraints on prosodic phrasing in children's speech**
Melissa A. Redford (University of Oregon), Zahra Foroughifar (University of Oregon), Laura C. Dilley (Michigan State University)

13:30 –13:50 **The role of lexical age-of-acquisition on phonetic variation in natural infant-directed speech**
Georgia Zellou (University of Pennsylvania), Rebecca Scarborough (University of Colorado), Eric Doty (University of Pennsylvania)

13:50–14:10 **Pre-babbling i nfants prefer lis tening t o in fant s peech: A l aunch pa d for t he perception-production l oop?** Matthew Masapollo (McGill University), Linda Polka (McGill University), Lucie Ménard (Université du Québec à Montréal)

14:10–14:30 **Baby s teps i n p erceiving a rticulatory foundations o f ph onological c ontrasts: i nfants detect audio→video congruency in native and nonnative consonants** Catherine T. Best (University of Western Sydney), Christian H. Kroos (University of Western Sydney), Sophie Gates (University of Western Sydney), Julia Irwin (University of Western Sydney)

14:30–16:20 **POSTER SESSION 1** (with coffee)

SESSION 4 Chair: Keiichi Tajima (Hosei University)

16:30 –16:50 **Phonetic vs. phonological factors in coronal-to-dorsal perceptual assimilation**
Eleanor Chodroff (Johns Hopkins University), Colin Wilson (Johns Hopkins University)

16:50 –17:10 **The usefulness of chaos: Lab versus non-lab speech for perceptual learning**
Elizabeth Casserly (Indiana University), David B. Pisoni (Indiana University)

17:10–17:30 **The effect o f p erceptual s imilarity in second la nguage le arning: P ositional as ymmetry i n phoneme substitution**
Yu-an Lu (National Chiao Tung University), Jiwon Hwang (Stony Brook University)

17:30 –17:50 **Sensitivity to fine acoustic detail affects comprehension of reduced speech in L2**
Ellen Aalders (Radboud University Nijmegen), Mirjam Ernestus (Max Planck Institute Nijmegen)

17:50 –18:10 **Voicing, F0, and phonological enhancement**
James Kirby (University of Edinburgh), D. Robert Ladd (University of Edinburgh)

18:10–18:40 **ALP GENERAL MEETING**

SATURDAY, July 26

SESSION 5 Chair: Takayuki Arai (Sophia University)

09:10 –10:00 **[Invited] Labeling in the wild: Crowdsourcing versus categorical perception**
Mark Hasegawa-Johnson (University of Illinois at Urbana-Champaign)

10:00 –10:30 **Comments and Discussion**: Natasha Warner (University of Arizona)

10:30 –10:50 **BREAK**

SESSION 6 Chair: Ian Wilson (University of Aizu)

10:50 –11:10 **The private life of stops: VOT in a real-time corpus of spontaneous Glaswegian**
Jane Stuart-Smith (University of Glasgow), Morgan Sonderegger (McGill University)
Rachel Macdonald (University of Glasgow), Thea Knowles (McGill University),
Tamara Rathcke (University of Kent)

11:10 –11:30 **Articulation rate and VOT in spontaneous speech**
Satsuki Nakai (University of Glasgow), James M. Scobbie (Queen Margaret University)

11:30 –11:50 **The syllable as a prosodic unit in Japanese lexical strata: Evidence from text-setting**
Rebecca L. Starr (National University of Singapore),
Stephanie S. Shih (University of California, Berkeley)

11:50 –12:10 **Lexical diffusion of vowel length merger in Seoul Korean: A corpus-based study**
Yoonjung Kang (University of Toronto Scarborough), Tae-Jin Yoon (Cheongju University),
Sungwoo Han (Inha University)

12:10 –13:10 **LUNCH / PUBLISHING IN LABORATORY PHONOLOGY: Q&A WITH THE EDITOR**

SESSION 7 Chair: Kiyoko Yoneyama (Daito Bunka University)

13:10–13:30 **The peril of sounding manly: A look at vocal characteristics of lawyers before the United States Supreme Court**
Alan C. L. Yu (University of Chicago), Daniel Chen (ETH Zurich), Katie Franich (University of Chicago), Jacob Phillips (University of Chicago), Betsy Pillion (University of Chicago), Yiding Hao (University of Chicago), Zhigang Yin (University of Chicago)

13:30 –13:50 **Stereotypes predict memory effects for voices**
Grant McGuire (University of California, Santa Cruz),
Molly Babel (University of British Columbia)

13:50 –14:10 **The role of exemplars in speech comprehension**
Annika Nijveld (Radboud University Nijmegen), Martijn Bentum (Radboud University Nijmegen), Louis ten Bosch (Radboud University Nijmegen)

14:10 –14:30 **Effects of phonological neighborhood density on phonetic variation: The curious case of French** Yao Yao (Hong Kong Polytechnic University), Christine Meunier (Aix Marseille Université)

14:30 –16:20 **POSTER SESSION 2** (with coffee)

SESSION 8 Chair: Shin-Ichiro Sano (Okayama Prefectural University)

16:30–17:20 **[Invited] Some cognitive factors behind vowel lengthening in spontaneous Japanese: A corpus-based study** Yasuharu Den (Chiba University)

17:20–17:50 **Comments and Discussion**: Shu-Chuan Tseng (Institute of Linguistics, Academia Sinica)

18:00– 20:00 **BANQUET** (and fireworks)

SUNDAY, July 27

SESSION 9 Chair: Shigeko Shinohara (Sophia University)

09:10 –10:00 **[Invited] On establishing the existence of word stress**
Carlos Gussenhoven (Radboud University Nijmegen)

10:00 –10:30 **Comments and Discussion**: Aditi Lahiri (University of Oxford)

10:30 –10:50 **BREAK**

SESSION 10 Chair: Shinichiro Ishihara (Goethe University Frankfurt am Main)

10:50 –11:10 **Pointed and plateau-shaped pitch accents in North Frisian dialects**
Oliver Niebuhr (Kiel University), Jarich Hoekstra (Kiel University)

11:10 –11:30 **Quantity contrast in Lule Saami: A three-way system**
Nora Fangel-Gustavson (CNRS/Sorbonne Nouvelle), Bruce Morén-Duolljá (University of Tromsø)

11:30 –11:50 **Prominence, phrasing, and information structure in Mawng (Australian)**
Janet Fletcher (University of Melbourne), Ruth Singer (University of Melbourne), Debbie Loakes (University of Melbourne)

11:50 –12:10 **No VOT perception without native VOT experience**
Rikke Bundgaard-Nielsen (La Trobe University), Brett Baker (University of Melbourne)

12:10 –13:10 **LUNCH**

13:10 –15:00 **POSTER SESSION 3** (with coffee)

SESSION 11 Chair: Mafuyu Kitahara (Waseda University)

15:10 –15:30 **Articulatory correlates of phonological relationships**
Kathleen Currie Hall (University of British Columbia), Hanna Smith (University of British Columbia), Kevin McMullin (University of British Columbia), Noriko Yamane (University of British Columbia), Blake Allen (University of British Columbia), Joash Gambarage (University of British Columbia)

15:30 –15:50 **Invariant coupling relations at the core of variable speech trajectories**
Leonardo Lancia (Max Planck Institute for Evolutionary Anthropology)

15:50–16:10 **Acoustic var iability ai ds t he i nterpretation of p honetic d etail i n c ross-language s peech production** Lisa Davidson (New York University), Sean Martin (New York University), Colin Wilson (Johns Hopkins University)

16:10–16:30 **General Discussion**

16:30–16:35 **Farewell**

MONDAY, July 28 (All Day)

Satellite Workshop

Theme: **Gestural coordination within and between speakers in first language phonological acquisition**

Organizers: Felicitas Kleber (Ludwig-Maximilian-Universität München)
Mary E. Beckman (Ohio State University)

LabPhon 14 POSTER SESSIONS

<u>FRIDAY, July 25</u> 14:30 –16:20 **POSTER SESSION 1** (with coffee)

P1-1 Comparing the discourse use of prosody in adolescents with normal hearing and cochlear implants
Colleen Holt (University of Melbourne), Ivan Yuen (Macquarie University),
David Rosson (Macquarie University), Katherine Demuth (Macquarie University)

P1-2 Code switch production in Cantonese-English bilingual children
Donald White (The Chinese University of Hong Kong), Peggy Mok (The Chinese University of Hong Kong)

P1-3 Modeling the acquisition of vowel normalization as cognitive manifold alignment
Andrew R. Plummer (The Ohio State University)

P1-4 Individual differences in children's prosodic focus-marking
Aoju Chen (Utrecht University/Max Planck Institute for Psycholinguistics)

P1-5 Effects of speaker language and listener language on children's stop place
Mary E. Beckman (Ohio State University, Columbus), Benjamin Munson (University of Minnesota, Twin Cities), Jan Edwards (University of Wisconsin, Madison)

P1-6 The acquisition of differential spectral kinematics of English sibilant fricatives
Patrick F. Reidy (The Ohio State University)

P1-7 Perception and production of phonemic vowel length in Australian English-learning 18-month-olds
Hui Chen (Macquarie University), Nan Xu Rattanasone (Macquarie University),
Felicity Cox (Macquarie University)

P1-8 A comparison of long consonant acquisition in Arabic, Finnish, Japanese and Welsh
Ghada Khattab (Newcastle University), Marilyn Vihman (University of York), Jalal Al-Tamami (Newcastle University), Satsuki Nakai (Glasgow University), Sari Kunnari (University of Oulu)

P1-9 Do infants normalize speaker and accent variability in vowel production?
Karen E. Mulak (University of Western Sydney), Samra Alispahic (University of Western Sydney)
Paola Escudero (University of Western Sydney)

P1-10 On lexical phonotactics and segmentability
Robert Daland (University of California, Los Angeles), Benjamin Börschinger (Macquire University)
Abdellah Fourtassi (Laboratoire de Sciences Cognitives et Psycholinguistique, ENS/EHESS/CNRS, Paris)

P1-11 Learning sound categories with phonotactics
Masaki Noguchi (University of British Columbia), Carla Hudson Kam (University of British Columbia)

P1-12 What kinds of units are used in speech perception?
Eva Reinisch (Ludwig Maximilian University Munich)

P1-13 Phonological and phonetic vowel reduction in two Ibero-Romance varieties
Marianna Nadeu (Penn State)

P1-14 The effect of English stress on non-native speakers' word recognition: Evidence from native speakers of lexical tone language Shu-chen Ou (National Sun Yat-sen University)

P1-15 Evidence of early motor planning in speech production
Donald Derrick (University of Canterbury/University of Western Sydney),
Romain Fiasson (University of Western Sydney)

P1-16 Acoustic effects of predictability and gender on Japanese high vowel reduction
James Whang (New York University)

P1-17 Tapping in American English: Context matters
Adam J. Chong (University of California, Los Angeles),
Megha Sundara (University of California, Los Angeles)

P1-18 Gestural coordination in word-initial Greek clusters as revealed by kinematic, articulatory, and acoustic data Jonathan Yip (The University of Hong Kong)

P1-19 Acoustic differentiation between coda reţroflex and denţal stops in Punjabi
Qandeel Hussain (Macquarie University), Ivan Yuen (Macquarie University)

P1-20 Stability in perceiving non-native segmental length contrasts
Yuki Asano (University of Konstanz), Bettina Braun (University of Konstanz)

P1-21 Effect of attention on L2 speech perception: Perception of the Korean three-way stop contrast by English learners Hyunjung Lee (University of Chicago)

P1-22 The perception of French reduced speech by native and non-native listeners
Sophie Brand (Radboud University Nijmegen),
Mirjam Ernestus (Radboud University Nijmegen/Max Planck Institute for Psycholinguistics)

P1-23 Subjective perception of affixation: A test case from Spanish

Anne Pycha (University of Wisconsin, Milwaukee)

P1-24 The psychological status of the right-branch condition and deaccentuation on *Sais-Sori* in Korean
Hyun Kyung Hwang (National Institute for Japanese Language and Linguistics)

P1-25 Perceptual relevance of Non-F0 acoustic correlates in Japanese accent
Yukiko Sugiyama (Keio University)

P1-26 The productivity and stability of competing generalizations in stress assignment
Paul Olejarczuk (University of Oregon)

P1-27 Effects of a sound change in progress on gender-marking cues in Japanese
Eun Jong Kong (Korea Aerospace University), Kiyoko Yoneyama (Daito Bunka University)
Mary E. Beckman (Ohio State University)

P1-28 Phonological Encoding and Articulation in the Absence of Metrical Spellout
Bryan B. Holbrook (University of California, Santa Cruz),
Alan H. Kawamoto (University of California, Santa Cruz)

P1-29 Perception-production asymmetry: The case of Hai-lu Hakka tone change and tone sandhi
Chia-Hsin Yeh (Michigan State University), Yen-Hwei Lin (Michigan State University)

P1-30 VC coarticulation in Taiwanese and its phonological consequences
Yueh-chin Chang (National Tsing Hua University), Hsieh Feng-fan (National Tsing Hua University)
Yi-cheng Chen (National Tsing Hua University)

P1-31 Posterior cavity and aperture distance oppositions in English coronal fricatives
Simon Gonzalez (The University of Newcastle), Mark Harvey (The University of Newcastle)
Michael Proctor (Macquarie University)

P1-32 Are acoustic cues sufficient for the identification and discrimination of Spanish approximants? Evidence from a production study
Mauricio Figueroa (University College London), Bronwen G. Evans (University College London)

P1-33 Acoustic and lingual variation within the 4-way contrast of vowel height in French
Laurianne Georgeton (CNRS/Univ. Paris 3-Sorbonne Nouvelle),Tanja Kocjančič Antolík (CNRS/Univ. Paris 3-Sorbonne Nouvelle), Cécile Fougeron (CNRS/Univ. Paris 3-Sorbonne Nouvelle)

P1-34 Syntactic predictability can facilitate the recognition of casually produced words in connected speech
Malte C. Viebahn (Max Planck Institute for Psycholinguistics),
James M. McQueen (Radboud University Nijmegen)

P1-35 Perceptual attunement to coarticulation: hearing tone in vowel height
Jason A. Shaw (University of Western Sydney), Wei-rong Chen (National Tsing Hua University),
Michael D. Tyler (University of Western Sydney), Donald Derrick (University of Western Sydney)
Michael Proctor (Macquarie University)

P1-36 Perceptual cues of Japanese /r/ Sounds: Formant transitions vs. intensity dip
Takayuki Arai (Sophia University)

P1-37 Prosodic strengthening on initial stops in English trochaic vs. iambic words
Jiseung Kim (University of Michigan), Sahyang Kim (Hongik University),
Taehong Cho (Hanyang University)

SATURDAY, July 26 14:30 –16:20 **POSTER SESSION 2** (with coffee)

P2-1 Rendaku in spontaneous speech
Shin-Ichiro Sano (Okayama Prefectural University)

P2-2 Catalan prepalatal allophony and phonological contrast
José I. Hualde (University of Illinois at Urbana-Champaign), Christopher Eager (University of Illinois at Urbana-Champaign), Marianna Nadeu (Penn State)

P2-3 Revisiting vowel harmony in Korean sound-symbolic words: a corpus study
Sang-Im Lee-Kim (New York University)

P2-4 Advancing corpus-based analyses of spontaneous speech: Switch to GECO!
Antje Schweitzer (Stuttgart University), Natalie Lewandowski (Stuttgart University),
Grzegorz Dogil (Stuttgart University)

P2-5 Phonetic and prosodic characteristics of disfluencies in French spontaneous speech
George Christodoulides (University of Louvain), Mathieu Avanzi (University Paris Diderot)

P2-6 Multilevel modelling of initial rise in French
Mathieu Avanzi (University Paris Diderot), George Christodoulides (University of Louvain)

P2-7 Stress-induced sibilant variations in Taiwan Mandarin spontaneous speech
Yu-Ying Chuang (National Taiwan University), Janice Fon (National Taiwan University)

P2-8 (t,d) deletion in everyday speech
Margaret E.L. Renwick (University of Georgia), Rosalind Temple (University of Oxford)
Ladan Baghai-Ravary (University of Oxford), John S. Coleman (University of Oxford)

P2-9 Using conversational corpus speech to investigate the impact of speech rate on speech perception performance over the adult life span
Xaver Koch (Radboud University Nijmegen), Esther Janse (Radboud University Nijmegen/Donders Institute for Brain Cognition, and Behavior/Max Plank Institute for Psycholinguistics)

P2-10 Speech reduction in Czech
Alice Kolman (Radboud University Nijmegen/Christian University of Applied Sciences CHE),
Petr Pollak (Czech Technical University)

P2-11 On the tail of the Scottish vowel length rule in Glasgow
Tamara Rathcke (University of Kent), Jane Stuart-Smith (University of Glasgow)

P2-12 Does /t/ produced as [ʔ] involve tongue tip raising? Articulatory evidence for the nature of phonological representations
Jennifer Heyward (The University of Edinburgh), Alice Turk (The University of Edinburgh),
Christian Geng (The University of Potsdam)

P2-13 A corpus based investigation of the contrast between French rise-fall and rise via wavelet based functional mixed models Cristel Portes (Aix-Marseille Université/CNRS/LPL),
Leonardo Lancia (Max Planck Institute for Evolutionary Anthropology)

P2-14 A new corpus of colloquial Korean and its applications
Kevin Tang (University College London), Brent de Chene (Waseda University)

P2-15 *Phonological CorpusTools*: A free, open-source tool for phonological analysis
Scott Mackie (University of British Columbia), Kathleen Currie Hall (University of British Columbia),
Blake Allen (University of British Columbia), Michael Fry (University of British Columbia),
Michael McAuliffe (University of British Columbia)

P2-16 Schwa reduction in spontaneous infant-directed speech
Mybeth Lahey (Max Planck Institute for Psycholinguistics/Radboud University Nijmegen/International Max Planck Research School for Language Sciences), Mirjam Ernestus (Radboud University Nijmegen/Max Planck Institute for Psycholinguistics)

P2-17 Uptalk in semi-spontaneous and scripted speech
Amanda Ritchart (University of California, San Diego), Amalia Arvaniti (University of Kent)

P2-18 Variability in the phonology of child-directed speech: evidence from a new naturalistic corpus
Sam J. Green (University College London)

P2-19 The perception of phrasal prominence in conversational speech
Jennifer Cole (University of Illinois), José I. Hualde (University of Illinois),
Tim Mahrt (University of Illinois), Christopher Eager (University of Illinois),
Suyeon Im (University of Illinois)

P2-20 The influence of prosodic boundaries on high vowel devoicing in Japanese
Oriana Kilbourn-Ceron (McGill University)

P2-21 Production of a non-phonemic contrast by native and non-native speakers: The case of American English flap Mafuyu Kitahara (Waseda University), Keiichi Tajima (Hosei University),
Kiyoko Yoneyama (Daito Bunka University)

P2-22 Trajectories of phonetic variability in spontaneous speech on reality TV
Morgan Sonderegger (McGill University)

P2-23 Toward a holistic measure of reduction in spontaneous speech
Michael McAuliffe (University of British Columbia), Molly Babel (University of British Columbia)

P2-24 Phonology Constrains the Distribution of the Particle *lah* in Singapore English
James Sneed German (Nanyang Technological University), Laurent Prévot (Aix-Marseille Université/Laboratoire Parole et Langage)

P2-25 Word-specific and sub-phonemic representations: yod-dropping and /u/-fronting in Derby
Márton Sóskuthy (University of York), Paul Foulkes (University of York),
Vincent Hughes (University of York), Jennifer Hay (University of Canterbury),
Bill Haddican (CUNY-Queens College)

P2-26 Verbal feedback: positioning and acoustics of French “ouais” and “oui”
Laurent Prévot (Aix-Marseille Université/CNRS/Laboratoire Parole et Langage), Jan Gorisch (Aix-Marseille Université/CNRS/Laboratoire Parole et Langage)

P2-27 Dissecting the consonant duration ratio
Erika Pillmeier (University of Potsdam), Stavroula Sotiropoulou (University of Potsdam),

Adamantios Gafos (University of Potsdam)

P2-28 Derived onsets in Spanish: an experimental study of resyllabification
Patrycja Strycharczuk (CASL-Queen Margaret University), Martin Kohlberger (Leiden University)

P2-29 Tashlhiyt syllabification: Perceptual evidence
Rachid Ridouane (CNRS/Univ. Paris 3-Sorbonne Nouvelle),
Pierre Hallé (CNRS/Univ. Paris 3-Sorbonne Nouvelle)

P2-30 Listener-specific perception of speaker-specific productions in intonation
Francesco Cangemi (University of Cologne), Martine Grice (University of Cologne)

P2-31 Focus intonation in Turkish
Canan Ipek (University of Southern California), Sun-Ah Jun (University of California, Los Angeles)

P2-32 Explicit and implicit gender priming in fricative perception
Benjamin Munson (University of Minnesota)

P2-33 Articulatory gestures and assimilation of mora nasal /ɴ/ in Japanese
Ai Mizoguchi (City University of New York), Douglas H. Whalen (Haskins Laboratories)

P2-34 Aspiration metathesis in Andalusian Spanish: The role of place of articulation
Hanna Ruch (University of Zurich)

P2-35 On /tl, dl/-misperception: Language specificity and cross-linguistic tendencies
Yuriko Matsumoto-Yokoe (Sophia University)

P2-36 Vowel harmony in French: Investigating formant trajectories
Agnieszka Duniec (Université de Nantes), Olivier Crouzet (Université de Nantes)

P2-37 Japanese nasal place/stricture assimilation: Electropalatographic evidence
Alexei Kochetov (University of Toronto)

<u>**SUNDAY, July 27**</u> 13:10 –15:00 **POSTER SESSION 3** (with coffee)

P3-1 Acoustic characteristics and variation of clicks in the endangered language N|uu
Carina Marquard (Kiel University), Oliver Niebuhr (Kiel University),
Alena Witzlack-Makarevich (Kiel University)

P3-2 Determining the representation of phonotactic restrictions with nonce words
Gillian Gallagher (New York University)

P3-3 "Chilcotin flattening" revisited: A phonetic investigation of Tsilhqut'in retraction effects
Sonya Bird (University of Victoria)

P3-4 Vowel-to-vowel coarticulation in Australian languages: Place matters
Simone Graetzer (University of Melbourne), Janet Fletcher (University of Melbourne),
John Hajek (University of Melbourne)

P3-5 Schwa in Tashlhiyt Berber in voiceless environments
Timo B. Roettger (University of Cologne), Rachid Ridouane (CNRS/Univ. Paris 3-Sorbonne Nouvelle),
Martine Grice (University of Cologne)

P3-6 Getting to the root of the problem: An ultrasound investigation of 'Advanced Tongue Root' in Lopit
Rosey Billington (University of Melbourne)

P3-7 Evidence of mismatch between tonal production and perception in Karbi
Amos Teo (University of Oregon), Linda Konnerth (University of Oregon)

P3-8 Inserted vowel articulation in Scottish Gaelic: a preliminary report
Diana Archangeli (University of Arizona /University of Hong Kong), Sam Johnston (University of Arizona), Jae-Hyun Sung (University of Arizona), Muriel Fisher (University of Arizona), Michael Hammond (University of Arizona), Andrew Carnie (University of Arizona)

P3-9 Tonal polarity in Xitsonga as a tonal artifact: an acoustic analysis
Siri Gjersøe (Humboldt University), Seunghun J. Lee (Central Connecticut State University)

P3-10 Is longer always better? Neurolinguistic evidence for asymmetric processing of consonant duration
Sandra Kotzor (University of Oxford), Adam Roberts (University of Oxford),
Allison Wetterlin (University of Oxford), Aditi Lahiri (University of Oxford)

P3-11 The articulation of Kaytetye coronals
Susan Lin (University of California, Berkeley), Mark Harvey (University of Newcastle),
Benjamin Davies (Macquarie University), Myfany Turpin (University of Queensland),
Alison Ross (Northern Territories Department of Education), Katherine Demuth (Macquarie University)

P3-12 Cross-dialectal vowel spaces of Greek
M. Baltazani (University of Ioannina/University of Oxford), E. Kainada (University of Ioannina),
K. Nikolaidis (Aristotle University of Thessaloniki), A. Sfakianaki (Aristotle University of Thessaloniki)

A. Lengeris (University of Crete/Aristotle University of Thessaloniki), E. Tsiartsoni (Aristotle University of Thessaloniki), D. Papazachariou (University of Patras), M. Giakoumelou (University of Patras)

P3-13 Laryngeal and tonal contrasts in the Tai dialect of Cao Bang
Pittayawat Pittayaporn (Chulalongkorn University), James Kirby (University of Edinburgh)

P3-14 Intra-oral pressure as independent of duration: Stops in Bininj Gun-wok
Hywel Stoakes (University of Melbourne)

P3-15 Testing the syllabicity of inserted vowels: A word-game experiment in Pnar
Hiram Ring (Nanyang Technological Institute), James Gruber (New Zealand Institute of Language Brain and Behavior), Jacqui Nokes (University of Canterbury)

P3-16 Wh-question intonation patterns in the Showamura dialect of Japanese
Jason Ginsburg (Osaka Kyoiku University), Ian Wilson (University of Aizu), Emiko Kaneko (University of Aizu), Naomi Ogasawara (University of Aizu)

P3-17 A study of the prosodic encoding of TOPICS as a category of information structure in Ngarinyman, a language of Australia
Candide Simard (SOAS), Faith Chiu (University College London), Eva Schultze-Berndt (University of Manchester)

P3-18 ERP correlates of two types of implicit knowledge of probabilistic phonotactics
Claire Moore-Cantwell (University of Massachusetts Amherst), Lisa Sanders (University of Massachusetts Amherst)

P3-19 An EMA examination of Czech post-aveolar and palatal consonants
Phil Howson (University of Toronto), Alexei Kochetov (University of Toronto)

P3-20 Inductive learning of long-distance dissimilation as a problem for phonology
Kevin McMullin (University of British Columbia), Gunnar Ólafur Hansson (University of British Columbia)

P3-21 ~~Asymmetries in cross-linguistic stop/fricative perception~~ cancelled
Youngja Nam (McGill University), Linda Polka (McGill University)

P3-22 Acoustic disjuncture in consonant clusters and vowel epenthesis
Suyeon Yun (Massachusetts Institute of Technology)

P3-23 L2 experience can hinder perception of non-native sounds
Jeffrey Holliday (Indiana University)

P3-24 Preferences in permitted sequences: A weighted markedness constraint model
Jeremy Perkins (University of Aizu)

P3-25 Effects of linguistic structure on perceptual attention given to different speech units
Shinae Kang (University of California, Berkeley), Keith Johnson (University of California, Berkeley)

P3-26 The perception of the tensity contrast in two German varieties
Conceição Cunha (Institute of Phonetics and Speech Processing, Munich), Ulrich Reubold (Institute of Phonetics and Speech Processing, Munich)

P3-27 Mismatch Negativity Reveals Abstract, Monovalent Phonological Primitives
Mirjam J.I. de Jonge (University of Amsterdam)

P3-28 Phonotactic learning and its interaction with speech segmentation
Alexis Black (University of British Columbia), Masaki Noguchi (University of British Columbia)

P3-29 Effects of a formant transition of preceding vowel off-glide on perception of Japanese geminate consonant *sokuon* Emi Yanagisawa (Sophia University), Takayuki Arai (Sophia University)

P3-30 An empirical assessment of the carryover bias in tonal coarticulation
Feng-fan Hsieh (National Tsing Hua University)

P3-31 Swedish focal accents and the syntax–prosody interface
Sara Myrberg (Stockholm University)

P3-32 Vowel timing in complex onsets of American English
Stavroula Sotiropoulou (University of Potsdam)

P3-33 The processing of schwa reduction in isolation and in connected speech: An electrophysiological Study Kimberley Mulder (Radboud University Nijmegen), Linda Drijvers (Radboud University Nijmegen)

P3-34 Applying phonological knowledge to phonetic accommodation
Emily Clare (University of Toronto)

P3-35 Realization of f0 peak alignment in Spanish by Korean L2 learners
Ji Young Kim (University of Illinois at Urbana-Champaign)

P3-36 The contribution of individual acoustic cues to the perception of focal prominence
Tim Mahrt (University of Illinois), Jennifer Cole (University of Illinois)

[DAY 1: Oral Session 1]

Infant-directed speech as a window into the dynamic nature of phonology

Reiko Mazuka[1,2], Andrew Martin[1], Yosuke Igarashi[3], Akira Utsugi[4]

1. RIKEN Brain Science Institute
2. Duke University
3. Hiroshima University
4. Nagoya University

Theoretical frameworks of phonology are built largely on the basis of idealized speech, typically recorded in a laboratory under static conditions. Natural speech, in contrast, occurs in a variety of communicative contexts where speakers and hearers dynamically adjust their speech to fit their needs. The present paper demonstrates that phonologically informed analysis of specialized speech registers, such as infant-directed speech, can reveal specific ways segmental and supra-segmental aspects of phonology are modulated dynamically to accommodate the specific communicative needs of speakers and hearers.

Data for the analyses comes from a corpus of Japanese infant-directed speech, consisting of 22 Japanese mothers' spontaneous speech directed to their infant child (Infant-directed speech, IDS) and an adult (adult-directed speech, ADS). The speech samples in the corpus are annotated with segmental, morphological and intonational information. We will show, for example, that the way intonation is exaggerated in Japanese IDS reflects the intonational structure of Japanese, which is different from that of English. We will also demonstrate that rules of phonological grammar, such as devoicing of high vowels and non-high vowels in Japanese, can be differently affected by the needs of the speaker to accommodate the specific characteristics of the listener.

[DAY 1: Oral Session 2]

Children's imitation of coarticulatory patterns in different prosodic contexts

Felicitas Kleber, Sandra Peters

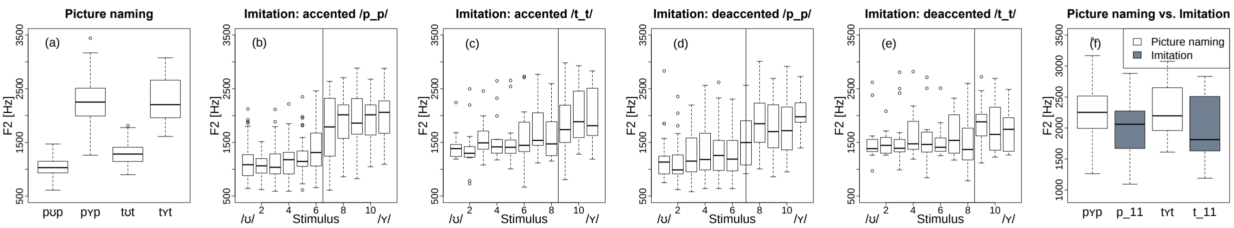
Institute of Phonetics and Speech Processing, Ludwig-Maximilians-University, Munich, Germany
{kleber|sandra}@phonetik.uni-muenchen.de

Acoustic speech signals show a great amount of synchronic variation that comes about through coarticulation, hypoarticulated speech – which is often observed in prosodically weak contexts (e.g. deaccented words) –, and other connected speech processes. Adult listeners usually compensate for coarticulation (Mann & Repp, 1980), but children have yet to learn to attribute acoustic modifications to the source that causes this synchronic variation. Nittrouer and Studdert-Kennedy (1987) showed that children compensated less for coarticulatory effects than adults; however, other studies reported that children overgeneralize adult forms (cf. Vihman, 1996). The aim of the present study was to investigate in an adult-child-interaction setting whether or not preschool children compensate for consonant-on-vowel coarticulation in prosodically weak and strong contexts. The overall aim is to relate these findings to Ohala's (1993) theory of sound change that sees unexperienced language users such as children as candidates for undercompensating for coarticulation and thus potentially initializing sound change.

In a picture naming task, 13 L1-German children aged 4;11 to 6;3 produced the names of puppets called /pʊp/, /pʏp/, /tʊt/, and /tʏt/ (five times each and always accented). While labial consonants have a retracting influence on vowels, alveolars have a fronting effect. The same children then imitated three times each 44 versions of the sentence *Maria hat* <target word> *gesagt* (*Maria said* <target word>) produced by a male adult speaker. The target word contained vowels from an 11-step continuum from /ʊ/ to /ʏ/ that was spliced into either a symmetric labial /p_p/ or an alveolar /t_t/ context. The only pitch accent was either on the target word (accented condition) or on *Maria* (deaccented condition). F2 was measured at the vowel's temporal midpoint of all target word productions and imitations. The analysis included only CVC tokens in which the children pronounced the consonantal context and prosody correctly.

An RM-ANOVA with F2 from the picture naming task as the dependent variable revealed a significant interaction effect for Context*Vowel ($F[1,2] = 13.4$, $p < 0.01$) showing coarticulatory effects in children's /tʊt/ productions ($t = -5.9$, $p < 0.001$) but not in /pʏp/ productions (cf. Fig. 1a). Comparatively high F2 values can also be seen in the imitations of stimuli 1–5 in the /t_t/-continua (cf. Fig. 1c+e). Irrespective of accentuation, children's imitations showed a steeper and earlier shift (indicated by the vertical lines in Figs. 1b–e) from back to front vowel realizations in /p_p/ than in /t_t/, suggesting that children attributed higher F2 values to the alveolar context and perceived and imitated more /ʊ/-like sounds in this context, i.e. they tended to compensate for coarticulation. Moreover, although /ʏ/ was realized with a high F2 and similar amount of variation in both contexts in production (median: /pʏp/ = 2253 Hz, /tʏt/ = 2197 Hz), the imitations of stimulus 11 from the accented alveolar continuum (i.e. an unambiguous /tʏt/ for adult listeners) showed vastly more variation in F2 and a significantly lower median (1813 Hz) than the /tʏt/ tokens from the picture naming task ($W = 846$, $p < 0.05$) as well as the labial stimulus 11 (2061 Hz; cf. Fig 1f which only contains accented tokens). The low F2 values for imitations of accented /t_t/ stimuli at the /ʏ/-end of the continuum imply that children tended to overcompensate for the contextual effect. In the deaccented condition, on the other hand, the shift from front to back vowels was more gradual (cf. Fig. 1d–e), i.e. children were less confident of the vowel categories leading to diminished compensation effects.

Figure 1 (a–f): F2 at the vowel's temporal midpoint separately for labial and alveolar, accented and deaccented conditions.



Children are sensitive to coarticulatory effects, but have yet to learn to compensate properly for these effects. Coarticulated sounds – particularly in prosodically weak position – are less stable in children's perception and may thus be more prone to undergo diachronic changes. Such changes are not necessarily due to undercompensation (Ohala, 1993): they may also come about due to overcompensation.

References

Mann, V.A., Repp, B.H. 1980. Influence of vocalic context on perception of the [ʃ]-[s] distinction. *Perception & Psychophysics* 28, 213-228. / Nittrouer, S., Studdert-Kennedy, M. 1987. The role of coarticulatory effects on the perception of fricatives by children and adults. *J. Speech and Hearing Research* 30, 319-329. / Ohala, J.J. 1993. The phonetics of sound change. In: Jones, C. (ed.): Historical Linguistics: Problems and Perspectives. London: Blackwell, 237-278. / Vihman, M. 1996. *Phonological Development*. Oxford: Blackwell.

An investigation into articulatory adaptation during acoustic mimicry of postvocalic /r/

Eleanor Lawson, James M. Scobbie, Jane Stuart-Smith

Queen Margaret University, University of Glasgow

elawson@qmu.ac.uk, jscobbie@qmu.ac.uk, Jane.Stuart-Smith@glasgow.ac.uk

Several decades of sociolinguistic investigation have established that there is an auditory dichotomy for postvocalic /r/ in the Scottish Central Belt, (Romaine 1978; Stuart-Smith 2003; Stuart-Smith 2007). Weak rhoticity is associated with working-class (WC) speech and strong rhoticity with middle-class (MC) speech.

Analysis of a socially-stratified ultrasound tongue imaging corpora of adolescent speech collected in the Scottish Central Belt has identified the articulatory variation that results in this auditory dichotomy; underlyingly coda /r/ in MC and WC speech involves radically different tongue shapes (Lawson, Stuart-Smith and Scobbie 2014) and tongue gesture timings (Lawson, Scobbie and Stuart-Smith, in preparation). This articulatory variation might be described as *covert*, as it has gone unidentified, despite decades of auditory and acoustic analysis; for example, it was always assumed that MC postvocalic /r/ was a retroflex or alveolar approximant (Romaine 1978; Speitel, Johnston 1983; Stuart-Smith 2003), yet UTI revealed bunched tongue configurations (Lawson et al 2014). Likewise, weak rhoticity in WC speech has been represented from an auditory perspective as various changes in prerhotic vowel quality, e.g. as pharyngealisation (Speitel & Johnston 1983), along with loss of the segmental rhotic consonant, yet UTI identified the presence of covert apical /r/ gestures with gesture maxima that occur after the offset of voicing or during the articulation of a following, masking consonant.

The fact that apparently covert articulatory variants pattern with speaker social class, suggests that covert articulatory variation in /r/ production, overlooked by sociolinguists and phoneticians, is nonetheless perceptible or recoverable from the audio. We present the preliminary results of a UTI-based, speech-mimicry study that investigates whether subtle articulatory variation in /r/ can be reproduced if the speaker is played an audio stimulus only.

As postvocalic /r/ lenition was most advanced in female speech in our most recently-collected corpus, we recruited six female Central-Belt Scottish speakers to take part in the mimicry study; three MC speakers aged 18-21 and three WC speakers aged 13-14. Baseline articulatory information on /r/ was gathered by recording audio and ultrasound while participants read a word list containing 23 (C)Vr(C) words e.g. *pore*, *farm*, *ear*, *fur* etc., plus 55 distractor words. All MC participants used bunched /r/ variants in baseline condition. Two of the WC participants used tip up or front up /r/ variants (see Lawson, Scobbie & Stuart-Smith 2011a) in baseline condition, the third used front up and front bunched variants (ibid.).

24 /r/-ful nonsense-word audio stimuli were extracted from the female speech section of a socially-stratified audio-ultrasound corpus of Glaswegian adolescent speech. Nonsense words were used to avoid potential effects of lexical access – a previous pilot study showed closer mimicry of utterances when the participant had not achieved lexical access (Lawson, Scobbie and Stuart-Smith 2011b). 12 working-class and 12 matching middle-class /r/-ful nonsense word stimuli were selected, where /r/ occurred after /i/, /a/, /ʌ/ and /o/ vowels. All WC audio stimuli had underlying tongue-tip raised /r/ articulations with delayed tongue-raising gestures. All MC stimuli had underlying bunched /r/ articulations with a lingual gesture that occurred early in the syllable. Stimuli were randomized with 58 nonsense-word distractors. Participants were asked to *mimic as closely as possible* the audio-only stimuli presented to them. Only one of the six participants reported that she thought /r/ was the focus of the study.

Analysis showed a range of /r/-mimicking behaviours from participants, including no modification in shape or timing from the baseline recording; altering the tongue shape from baseline, but not matching the stimulus and altering the tongue shape from baseline to match the stimulus. One speaker was particularly good at matching the underlying tongue shape associated with the audio stimulus and also adjusted her tongue gesture timings much more than the other speakers in order to match the underlying tongue gesture timings of the stimuli. Some weakly /r/-ful stimuli were mimicked as /r/-less utterances; however, the fact that the majority of the weakly /r/-ful stimuli were copied as /r/-ful utterances, suggests that cues remain in the audio signal that convey the presence of an underlying /r/ gesture.

References

Lawson, E., Scobbie, J.M. and Stuart-Smith, J., 2011a. The social stratification of tongue shape for postvocalic /r/ in Scottish English1. *Journal of Sociolinguistics,* 15(2), pp. 256-268. / Lawson, E., Scobbie, J.M. and Stuart-Smith, J., 2011b. A single-case study of articulatory adaptation during acoustic mimicry, *Proc. 17th ICPhS*, 17th - 21st Hong Kong, 1170-1173. / Romaine, S., 1978. Postvocalic /r/ in Scottish English: Sound change in progress. In: P. TRUDGILL, ed, *Sociolinguistic Patterns in British English.* pp. 144-158. / Speitel, H.H. and Johnston, P.A., 1983. *A Sociolinguistic Investigation of Edinburgh Speech.* End of Grant Report. Economic and Social Research Council. / Stuart-Smith, J., 2007. A sociophonetic investigation of postvocalic /r/ in Glaswegian adolescents, *Proc. 16th ICPhS*, Saarbrucken, 1307. / Stuart-Smith, J., 2003. The phonology of modern urban Scots. In: J. Corbett, J.D. McClure and J. Stuart-Smith, eds, *The Edinburgh Companion to Scots.* 1 edn. Edinburgh, U.K.: Edinburgh University Press, pp. 110-137.

Dolls are prissy: Preschoolers' toy preferences predict medial /t/ production

Jacqui Nokes
University of Canterbury, Christchurch, New Zealand
jacqueline.nokes@canterbury.ac.nz

Phonetic research in L1 acquisition typically studies children's speech to determine when and how children come to correctly produce the sounds of their native language. The definition of “correct” is usually tacit, assuming an acceptable range of linguistic variation without definition or scope; the role that social variation has in determining a child's (target) phonetic repetoire is seldom considered (Munson, 2009). This paper investigates the relationship between the development of gender identity and sex-correlated phonetic variation. Children begin to express their social group membership by showing preference for same-sex peers, and sex-stereotypical clothing, activities and toys in the late preschool years (Martin and Little, 1990). There is evidence that they also begin to moderate their speech to reflect in-group norms around the same age (Foulkes and Docherty, 2006). This paper presents the first results from ongoing research exploring the emergence of gender-based sociophonetic variation amongst New Zealand pre-schoolers.

Samples of speech from New Zealand English-speaking boys and girls aged 3 - 5 years were recorded during a picture naming task, and from natural conversation during activities with an interviewer and parents. Parents were also recorded producing the words from the picture naming task. Gender identity development was indexed using methodology drawn from social psychology (Zucker, 2005). Children were questioned on their toy and peer preferences, asked to categorise sex-stereotypical clothing and toys, and discussed sex-category identification and membership. In particular, children were asked to select their favourite toys from a collection of items, where three were stereotypical girls' toys (doll, pony, purse), three were stereotypical boys' toys (car, helicopter, snake), and three were neutral (Etch-a-Sketch, bear, kaleidoscope). The children were later asked to categorise the toys according to who liked to play with them (girls, boys or both).

New Zealand English has an incoming fricated variant of medial-/t/ that is more prevalent in the speech of young women than men. A total of 810 tokens containing medial-/t/ were extracted from the speech recordings of 51 children, inspected for their acoustic characteristics in Praat, and coded auditorily for closure, voicing and frication. In line with previous research (Fiasson, 2013), there was a greater incidence of fricated medial-/t/ in speech drawn from the mothers than the fathers in this study; and this was also reflected in speech tokens drawn from the girls and boys.

Mixed model regression was used to identify factors affecting the frication of medial /t/. In a model including traditional sociolinguistic predictors age and sex of the child, mother's socioeconomic index (MSEI) and word frequency, only mother's socioeconomic index and an interaction between sex and MSEI reached or came close to significance; de facto R^2 values (Nakagawa and Schielzeth, 2013) were 0.047 for the fixed effects, and 0.319 for the complete model.

In an alternative model, a measure indicating toy preference (0 for girls' toys, 1 for neutral and 2 for boys' toys) was added. In this model, toy preference replaced all other variables as the sole predictor, with greater significance. Estimated R^2 for fixed effects improved to 0.055 and for the full model to 0.323.

Figure 1: Alternative models predicting medial /t/ frication by peschoolers:
using traditional measures (left) and using toy preference (right)

	Estimate	**Std Error**	**Z value**	**Pr(>|z|)**
(Intercept)	-3.234	1.050	-3.080	0.002
GenderM	2.417	1.417	1.705	0.088
MSEI	0.036	0.017	2.203	0.028
GenderM:MSEI	-0.045	0.023	-1.952	0.051

	Estimate	**Std Error**	**Z value**	**Pr(>|z|)**
(Intercept)	-0.614	0.288	-2.133	0.033
TPmode	-0.531	0.183	-2.904	0.003

The results suggest that burgeoning awareness of social group membership and the identification of extra-linguistic in-group markers are reflected in the adoption of gendered social variation within speech. Furthermore, qualitative measures of social group membership that allow for variation in gender expression may have more explanatory value in modelling social variation than the use of broad descriptive categories such as sex.

References

Fiasson, R. 2013. *Allophonic Imitation Within and Across Word Positions*. Unpublished doctoral dissertation, University of Canterbury, Christchurch, New Zealand. / Foulkes, P., and Docherty, G. 2006. The social life of phonetics and phonology. *Journal of Phonetics* 34:3, 409-439. / Martin, C. L., and Little, J. K. 1990. The Relation of Gender Understanding to Children's Sex-Typed Preferences and Gender Stereotypes. *Child Development,* 61, 1427-1439 / Munson, B. 2009. Pathology or social indexing? In C. Bowen (Ed.), *Children's Speech Sound Disorders*. Oxford: Wiley-Blackwell, 342-346. / Nakagawa, S. and Schielzeth, H. 2013. A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution* 4, 133–142. / Zucker, K. J. (2005). Measurement of Psychosexual Differentiation. *Archives of Sexual Behavior* 34, 375-389.

[DAY 1: Oral Session 3]

Constraints on Prosodic Phrasing in Children's Speech

Melissa A. Redford[1], Zahra Foroughifar[1], Laura C. Dilley[2]
[1]University of Oregon and [2]Michigan State University
redford@uoregon.edu, zforough@uoregon.edu, ldilley@msu.edu

Prosody mediates between language and speech in production, defining the realm of speech planning (Keating & Shattuck-Hufnagel, 2002; Krivokapić, 2012). Prosodic structure, especially phrasing, is strongly influenced by syntax (Price et al., 1991); but non-linguistic factors, such as working memory and speech rate, are also important (Choe, 2013; Jun, 2003). Here we address the relative influence of these linguistic and non-linguistic factors on boundary placement and phrase length in children's speech across different age groups and over developmental time. The goal is to illuminate constraints that best explain individual and age-related differences in the realization of prosodic structure.

Structured spontaneous speech was recorded from 70 American-English speaking children once a year for 3 years. The current analyses investigate how changes in syntactic complexity, working memory capacity, and speech rate effected children's prosodic phrasing across years. The predictions were that age-related differences and developmental changes in boundary placement would be best explained by syntactic complexity, but that changes in phrase length would be better explained by the non-linguistic factors. These predictions were motivated by the hypothesis that non-linguistic factors exert temporal constraints on speech planning that in turn condition message chunking, which is meaning-bound. To test the hypothesis, prosodic phrase boundaries were identified by a team of analysts in 30 seconds samples drawn from each of the 210 recordings. A formal prosodic transcription system, the RaP System (Breen et al., 2012), was used. Analysts also noted disfluencies and made a metalinguistic judgment regarding the appropriateness of a boundary given the semantic-syntactic context. This judgment is our measure of boundary location. Going forward, boundary location will be defined with respect to syntactic constituency. As for the predictor variables, syntactic complexity was coded independently. The average number of phrasal nodes per clause (calculated for all Year 1 samples) correlated highly with the simpler measure of average clause length in words (r = .89, p < .001), which we use here. The measure of working memory was the average raw score from standardized repetition recall tasks administered yearly. Speech rate, measured in syllables per second (pause inclusive), correlated highly with articulation rate (r = .78, p < .001), and so likely captures aspects of motor skill development (Anonymous, submitted).

Children were grouped by age for the analyses. A younger age group included children who were 5 and 6 years old in Year 1; an older group, 7 and 8 years old. Study Year was a repeated measures factor. Subject was a random factor. GLMM analyses indicated significant increases across study years in the appropriateness of boundaries [F(2, 203) = 3.67, p = .027] and in prosodic phrase length [F(2, 206) = 13.57, p < .001], which was measured in syllables. Age Group also had a significant effect on phrase length [F(1, 206) = 5.97, p = .015] and interacted with the effect of Study Year for boundary appropriateness [F(2, 203) = 3.09, p = .048]. When syntactic complexity, working memory, and speech rate were added as predictors, with the ratio of disfluencies-to-boundaries added as a control, only the control variable was significant for boundary appropriateness [F(1, 185) = 68.33, p < .001], suggesting that boundary placement in fluent speech was consistent with semantic-syntactic structure regardless of a child's age. This result is consistent with chunking as meaning-bound. On-going work will determine whether syntactic complexity might account for any developmentally-related changes in boundary placement at stronger and weaker syntactic junctures. With regard to prosodic phrase length, age-related and individual differences in length were predicted by speech rate [F(1, 187) = 52.06, p < .001] and by the control variable [F(1, 187) = 21.07, p < .001]. The effect of syntactic complexity was weaker [F(1, 187) = 5.20, p = .024], and not significantly correlated with phrase length on its own. However, in the context of the other variables, greater syntactic complexity was associated with shorter phrase lengths. This result suggests an interaction between language and speech planning; perhaps a prosodic strategy for coping with complexity. As for the strong effect of speech rate on phrase length, this was likely due to a related finding: average phrase duration, measured in milliseconds, was constant across Age Group and Study Year. Together, the length and duration results are consistent with non-linguistic temporal constraints on planning. The absence of a working memory effect on phrase length may call into question the adequacy of the particular measure used in the current study.

References

Anonymous. The perceived clarity of children's speech varies as a function of their default articulation rate. Submitted / Breen, M., Dilley, L., Kraemer, J., & Gibson, E. (2012). Inter-transcriber agreement for two systems of prosodic annotation: ToBI (Tones and Break Indices) and RaP (Rhythm and Pitch). *Corpus Linguist. Linguist. Theory*, 8, 277-312. / Choe, W.-K. (2013). *The role of supralexical prosodic units in speech production.* Unpublished PhD thesis, Univ. of Oregon, Eugene. / Jun, S.-A. 2003. The effect of phrase length and speech rate on prosodic phrase length. *Proc. 15th ICPhS Barcelona*, 483-486. / Keating, P. & Shattuck-Hufnagel, S. (2002). A prosodic view of word form encoding for speech production. *UCLA working papers in phonetics*, 101, 112-156. / Krivokapić, J. (2012). Prosodic planning in speech production. In S. Fuchs et al. (Eds.) *Speech planning and dynamics* (pp. 157-190). Frankfurt, Germany: Peter Lang. / Price, P. J., Ostendorf, M., Shattuck -Hufnagel, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. *J. Acoust. Soc. Am.*, 90, 2956-2970.

The role of lexical age-of-acquisition on phonetic variation in natural infant-directed speech

Georgia Zellou, Rebecca Scarborough, Eric Doty

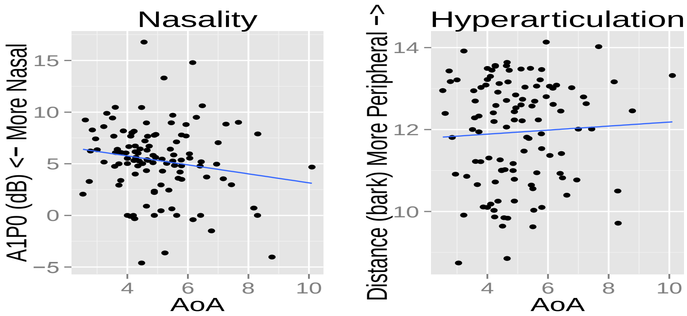
University of Pennsylvania, University of Colorado, University of Pennsylvania

gzellou@sas.upenn.edu, rebecca.scarborough@colorado.edu, ericdoty@sas.upenn.edu

This study investigates the relationship between fine-grained phonetic variation and word properties in natural speech directed toward infants (IDS). In speech directed toward adult interlocutors, both from spontaneous conversations and in controlled laboratory studies, two notable lexical variables—word frequency and phonological neighborhood density—have systematic effects on phonetic realization. Low frequency and high neighborhood density words consistently show an expanded vowel space relative to high frequency and low neighborhood density words (e.g., Wright, 2004), as well as increased coarticulation (Scarborough, 2013). These factors affect lexical perception as well. Low frequency and high ND words elicit slower and less accurate responses in auditory perception tasks due to lower activation and greater competition in lexical access. In fact, it has been suggested that the production patterns arise as attempts on the part of the speaker to compensate for the perceptual difficulty for a listener that might be associated with a particular word. Since mothers talking to their pre-linguistic infants have been shown to structure their speech to be maximally intelligible, we might expect to see such lexically-based listener-directed phenomena, where difficult words are produced in a perceptually-enhanced manner, in IDS. However, infants have such small amounts of linguistic exposure, relative to adults, and such small lexicons, which do not yet exhibit the organization of a more mature lexicon. Therefore, frequency and neighborhood density do not seem like relevant factors in assessing lexical difficulty for infants. In this study, we consider Age-of-Acquisition (AoA) (Kuperman, et al., 2012), or the reported age at which a word is learned, as a metric of lexical difficulty. Shorter perception and production latencies of earlier acquired words, above and beyond effects of predictable lexical variables, have been reported both in adults (Brysbaert et al., 2011) and in children preschool-aged and older (Garlock et al., 2001). We predicted that AoA would condition variation in phonetic variation as well, in particular, in IDS.

To test these predictions, we extracted all monomorphemic content words containing a nasal segment (either pre-vocalic or post-vocalic) from the Brent corpus (Brent & Siskind, 2001) of 16 mothers recorded in their homes talking to their infants (age 9-16 mo.). Vowel nasality (difference between nasal and oral formant spectral peak amplitudes, = A1-P0, Chen 1997) and Euclidean distance from vowel space center (in bark) were calculated at vowel midpoint. The data were fitted to linear mixed models with predictors of AoA, frequency-weighted neighborhood density, and log frequency. Collinearity of predictors (particularly, frequency with other predictors) was handled through residualization. Random intercepts of speaker and item were included. There was a significant effect of AoA in both models: later-acquired words were produced with both greater vowel hyperarticulation [estimate=1.26, t=1.74, p=0.008] and greater vowel nasalization [estimate=-0.06, t=-2.58, p=0.02] than early-acquired words, as shown in Figure 1.

Figure 1: Vowel Nasality (left) and Hyperarticulation by AoA measured in 8,127 tokens of 109 words items from the Brent corpus



These results demonstrate (1) that AoA is a predictor of effects in production, as well as in perception, and (2) that infant directed speech accommodations take lexical difficulty, measured in terms of AoA into account.

References

Brent, M. R. & Siskind, J. M. 2001. The role of exposure to isolated words in early vocabulary development. *Cognition*, 81, B33-B44. / Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A.M., Bölte, J., & Böhl, A. (2011). The word frequency effect: A review of recent developments and implications for the choice of frequency estimates in German. Experimental Psychology, 58, 412-424./ Chen, M. 1997. Acoustic correlates of English and French nasalized vowels. *J. Acoust. Soc. of Am.*, 102, 2360-2370. / Garlock, V.M., Walley, A.C., & Metsala, J.L. 2001. Age-of-acquisition, word frequency, and neighborhood density effets on spoken word recognition by children and adults. *JML*, 45, 468-492. / Kuperman, V., Stadthagen-Gonzalez, H., Brysbaert, M. 2012. Age-of-acquisition ratings for 30,000 English words. *Beh. Res. Methods*, 44, 978-990. / Scarborough, R. 2013. Neighborhood-conditioned patterns in phonetic detail: relating coarticulation and hyperarticulation. *J. Phonetics*, 41, 491-508. / Wright, R. 2004. Factors in lexical competition in vowel articulation. In J. Local, R. Ogden & R. Temple (eds.), *Papers in Laboratory Phonology VI* (pp. 26-50). Cambridge Univ. Press.

Pre-babbling infants prefer listening to infant speech: A launch pad for the perception-production loop?

Matthew Masapollo[1], Linda Polka[1] & Lucie Ménard[2]

[1]McGill University, [2]Université du Québec à Montréal

matthew.masapollo@mail.mcgill.ca; linda.polka@mcgill.ca; menard.lucie@uqam.ca

The interplay between the development of early speech perception and production capacities is not yet fully understood. One unexplored aspect of this topic is infants' perception of the segments and syllables they *themselves* produce. For infants to learn to produce intelligible words, they must effectively monitor and assess their own *self-produced* speech, which entails perceiving speech produced by an *infant* speaker (Kuhl & Meltzoff, 1996; Anonymous, submitted). Yet, very little is known about how young infants respond to infant-produced speech. Other research shows that infants acquiring different languages develop babbling with distinct vowel characteristics when measured acoustically (Rvachew, Mattock & Polka, 2006). Such findings suggest that infants actively engage in the processing of their *own* vocal output to link sensory patterns that they have heard with sensorimotor patterns that they are attempting to imitate. A listening bias favoring speech with infant vocal properties may contribute to infants' precocious vocal learning.

Here, we provide the first evidence that *pre-babbling* infants display a listening preference for infant speech over adult speech, and begin to characterize the properties of infant speech that capture infants' attention. Across four experiments, infants between the ages of 3 and 6-months were tested in a sequential preferential listening procedure. In this procedure, infants were seated on their caregiver's lap and shown a static checkerboard pattern to look at while listening to isolated /i/ vowels synthesized to emulate productions by adult female and infant speakers. The vowels were synthesized using the Variable Linear Articulatory Model (VLAM; see Ménard, Schwartz & Boë, 2004). Infants were presented with the infant vowels and adult vowels on alternating trials. Trial length was infant-controlled, i.e., trials were initiated by infant fixation on a static visual pattern, and ended when infants looked away for more than 2 seconds. In each experiment, we collected 12 trials, 6 for each speaker type, and then calculated whether infants choose to listen longer to infant vowels over adult vowels. The type of speaker that infants heard first was counterbalanced in each experiment.

In Experiment 1, infants listened significantly longer to infant vowels than to adult female vowels, when both voice pitch and vocal resonances were age-appropriate (M_{Infant}=10.9 sec, *SD*=3.5; M_{adult}=8.1, *SD*=2.2), $t(17)=4.051$, $p=.001$, $r^2=.70$. However, in Experiment 2, infants failed to show a listening preference when age-appropriate vocal resonances were maintained (adult vs. infant) but voice pitch values were matched to conform to pitch values typical of infant speakers (315, 360 Hz), $t(15)=1.003$, $p=.332$. Likewise, in Experiment 3, infants also displayed no evidence of a pitch-related listening preference when presented vowels with infant resonance properties produced with a higher (450 Hz) versus a lower (360 Hz) voice pitch, $t(17)=-.252$, $p=.804$. Thus, infants' bias for the infant vowels cannot be attributed to a preference for higher pitch *per se*; rather, infants appear to prefer speech with voice pitch values falling within the natural range of an infant talker, which interestingly, also overlaps with pitch values observed in infant-directed speech.

In Experiment 4, we investigated whether infants' preference for infant vowels extends beyond pitch information. In particular, we tested whether infants prefer infant vowels over adult vowels when both types of vowels are synthesized with pitch values characteristic of a female speaker (210 or 240 Hz). If infants show a preference for the infant vowels here, it would suggest that infants are *also* attracted to vocal resonance properties that specify an infant talker. Indeed, the results revealed a significant difference in infants' listening times such that they listened longer to low pitch vowels with infant resonance properties (M_{Infant}=8.1 sec, *SD*=2.8; M_{adult}=6.8, *SD*=2.0), $t(21)=2.216$, $p=.038$, $r^2=.435$.

These results suggest that young infants exhibit a perceptual bias for speech signals with infant-like voice pitch and vocal resonance. Such a bias may motivate infants to engage in babbling and facilitate infants' processing of their own self-produced auditory feedback during phonological development, which in turn, might help to initiate the perception-production loop. Additionally, these new findings have implications for our theoretical perspective on the role of increased pitch in infant-directed speaking styles.

References

Anonymous. Who's talking now: Infants' perception of vowels with infant vocal properties. Submitted. / Kuhl, P.K., Meltzoff, A.N. 1996. Infant vocalizations in response to speech: Vocal imitation and developmental change. *J. Acoust. Soc. Am.* 100, 425-438. / Ménard, L., Schwartz, J-L., Boë, L.-J. 2004. The role of vocal tract morphology in speech development: Perceptual targets and sensori-motor maps for French synthesized vowels from birth to adulthood. *J. Speech Lang. Hear. Res.* 47, 1059-1080. / Rvachew, S., Mattock, K., Polka, L. 2006. Developmental and cross-linguistic variation in the infant vowel space: The case of Canadian English and Canadian French. *J. Acoust. Soc. Am.* 120, 2250-2259.

Baby steps in perceiving articulatory foundations of phonological contrasts: Infants detect audio→video congruency in native and nonnative consonants

Catherine T. Best[1,2,3], Christian H. Kroos[1], Sophie Gates[1] and Julia Irwin[2,4]

[1]MARCS Institute, U Western Sydney, Australia, [2]School of Humanities & Communication Arts, U Western Sydney, Australia, [3]Haskins Laboratories, USA, [4]Southern Connecticut State U, USA

c.best@uws.edu.au chkroos@gmail.com s.gates@uws.edu.au irwinj1@southernct.edu

Theoretical background. Whether the perceptual primitives that perceivers use to recognize phonological distinctions include articulatory information or, as widely assumed, only acoustic information remains an important topic of debate. The answer is central to understanding infants' early steps toward acquisition of their native phonology. Compatible with the premise that perceivers detect articulatory information in speech, adults' perception of acoustic speech is systematically influenced by visible speech movements: a synchronously-presented phonetically-congruent talking face enhances intelligibility of speech in noise, whereas a phonetically-incongruent face can produce "misperception" of the audio token's place of articulation (McGurk effect). Similar results have been found in young infants. But bimodal audio-visual (AV) studies alone cannot pinpoint whether perceivers detect amodal *articulatory correspondences* between audio and video signals, or instead rely on *learned AV associations* between specific audio speech categories and the co-occurring face motions they have experienced with those sounds. Resolving this issue requires examination of: 1) *intermodal* recognition (cross-modal "transfer") of the common information between audio-only (A) and video-only (V) tokens of a given speech target; and of 2) developmental changes in recognition of experienced (native) versus non-experienced (nonnative) speech contrasts. The only prior such study tested English vs. Spanish infants' detection of A→V congruence in a visible contrast used in English but not Spanish: /b/-/v/. Both groups succeeded at 4 months but only English infants did so at 11 months (Pons, Lewkowicz, Sebastián-Gallés & Soto-Faraco, 2009). The 4-month results refute the hypothesis that perception of intermodal congruence rests on learned A-V associations. However, because both stimuli were labials (same articulator: LIPS) we can't tell whether they detected AV congruence based on articulator-distinct or other information (e.g., phonetic features; cross-modal psychophysical properties). Moreover, as Spanish 11-month-olds fail to discriminate audio-only /b/-/v/, their lack of differential response to the silent /b/ vs. /v/ videos after being habituated to audio /b/ or /v/ tells us nothing about their ability to detect intermodal congruencies. Therefore, we designed a series of studies on 4- and 11-month-old infants' detection of intermodal A→V congruency for visibly-distinct between-articulator contrasts (LIPS-TONGUE TIP) in native and two types of nonnative consonants.

Experiments. Three experiments assessed sensitivity to intermodal A→V congruence in Australian English-learning infants of 4 months (n = 20/expt) and 11 months (n = 20/expt). The audio-only stimuli used for the habituation phase of the task were natural tokens produced by native female speakers of each target language: English stops /pa/-/ta/ (Expt 1); Tigrinya ejectives /p'a/-/t'a/ (Expt 2); !Xóõ bilabial vs. dental clicks /ʘa/-/ǀa/ (Expt 3). The Tigrinya contrast was chosen because it has been shown that English adults and infants discriminate the audio contrast easily, but that adults nonetheless clearly hear these items as non-English-like (e.g., "choked" or "coughed" deviant variants of P vs. T). The Xóõ clicks were chosen for comparison, as prior findings indicate English adults hear them as nonspeech articulations ("kiss" vs. "tsk" sounds). In a separate study with Australian 4- (n = 30) and 11-month-olds (n = 25) we confirmed that both ages can discriminate audio-only tokens of these clicks. The silent-video stimuli in the A→V experiments were of an American female producing English /ba/ and /da/, which match the place of articulation difference for each target contrast, yet use a different speaker and voicing feature all three audio contrasts. The A→V conditioned visual fixation task had four phases: 1) habituation to the bilabial or coronal audio item (with on-screen checkerboard display; 2 subgroups/age of n = 10); 2) two "lag" trials of additional audio target presentations to assure habituation was maintained; 3) two video-only "buffer" trials of the video talker blinking her eyes, to familiarize infants to the switch to silent video stimuli; 4) test phase in which two trials presented the silent-video bilabial vs. coronal articulations. Intermodal sensitivity was evaluated by comparing infants' fixation times to the video trials that were articulator-congruent vs. articulator–incongruent to the infant's audio habituation stimulus, via ANOVAs at each age/experiment.

Results and Discussion. Four-month-olds showed a fixation preference for A→V articulator congruency in both the English and Tigrinya contrasts, but 11-month-olds only preferred A→V congruency for English, reversing to a novelty preference, i.e., for A→V *in*congruency, with the Tigrinya ejectives. In striking contrast, the 4-month-olds showed an A→V *in*congruency preference for the !Xóõ clicks, while the 11-month-olds failed to detect any A→V congruency relationship. We propose that both 4- and 11-month-olds recognize the Tigrinya ejectives as speech articulations, but only the older group recognize that they are deviant from English. In contrast, 4-month-olds already recognize !Xóõ clicks as deviant from English, while 11-month-olds instead treat them as nonspeech, like adults. We will discuss the implications of these findings for theories of native-language perceptual attunement and phonological development.

[DAY 1: Oral Session 4]

Phonetic vs. phonological factors in coronal-to-dorsal perceptual assimilation

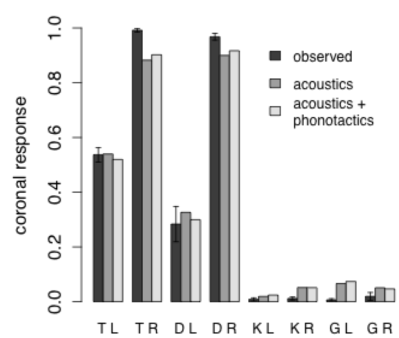
Eleanor Chodroff and Colin Wilson
Johns Hopkins University
chodroff@cogsci.jhu.edu, wilson@cogsci.jhu.edu

A central issue in cross-language speech perception is the relative influence of phonetic and phonological knowledge of the native language (e.g., Best & Strange, 1992). We address this issue by studying English listeners' perception of Hebrew stop-liquid clusters. Previous research has demonstrated that English (and French) listeners often identify Hebrew word-initial /tl/ and /dl/ clusters as beginning with dorsals (/k/ and /g/; Hallé & Best, 2007). While the phonological ban on initial coronal-lateral clusters in English may partially account for this misperception, it is also possible that pre-lateral Hebrew coronal stops are perceptually assimilated to English dorsal stops on acoustic-phonetic grounds (Flemming, 2007). We investigated whether phonetic properties of Hebrew stop bursts and transitions can account for English listeners' perceptions, and the extent to which incorporating phonotactic constraints improves this account.

Stimuli were stop-liquid-vowel syllables (4 recordings each of /t,d,k,g/ x /l,ʁ/ x /a,e,i,o,u/) produced by a female native speaker of Israeli Hebrew in the frame *Tagit __ shuv* ('Say __ again'). Prompts for the recordings were presented in Hebrew orthography with supporting IPA transcriptions. English listeners (N=18) identified the initial consonant of each syllable in an unspeeded forced-choice task with all six English stops as response options. In each trial, a single syllable (excised from the frame) was presented twice consecutively; syllable order was randomized separately for each participant. The results display a clear perceptual assimilation effect, with pre-lateral coronal stops frequently perceived as dorsals; coronals before the uvular approximant and dorsals in all environments were identified much more accurately. A Bayesian mixed-effects logistic model with random effects for participants and items verified the significant interaction between stop place and following context on identification accuracy ($p < .001$). Although /tl/ was identified more accurately overall than /dl/ (cf. Hallé & Best, 2007), this effect was not systematic across tokens. The interaction of stop place and following consonant remained significant when vowel category was included in the analysis.

Several acoustic properties known to cue stop place in English were measured in the Hebrew stimuli: spectral center of gravity (COG) of the initial 10 ms burst transient (measured with a smoothed spectrum to minimize variability); peak amplitude of the transient relative to the maximum of the following liquid; positive voice onset time; and the difference in Hz between F3 and F2 at the onset of voiced formant structure. A logistic model with these acoustic-phonetic predictors alone provided a close match to the observed rate of *coronal responses* across all legal and illegal cluster types (AIC = 1386, BIC = 1421; Fig.1). Consistent with English phonetic patterns, stops were more likely to be identified as coronal if they had higher COG values ($\beta = 3.14$, $p < .001$), greater burst amplitude ($\beta = 0.28$, $p < .001$), shorter VOT ($\beta = -0.63$, $p < .001$), or a larger F3-F2 difference ($\beta = 0.28$, $p < .001$). Because all predictors were scaled prior to analysis, the results suggest that burst COG was the primary cue used by English listeners to distinguish coronal from dorsal place in this experiment. A more complex model containing all of the acoustic-phonetic predictors and a phonotactic constraint against illegal clusters (/tl/, /dl/) showed a better fit to the data (AIC = 1335, BIC = 1389). However, the phonotactic factor was non-significant ($p = .09$) and its effect on the predicted coronal identifications was minimal.

Figure 1: Observed and predicted coronal responses for each consonant cluster.



Many theories of perceptual assimilation assume that phonotactic constraints force non-native structures to be mapped to phonetically similar native categories (Dupoux et al., 2011). The present results suggest that phonetic similarity alone can suffice to account for the type and rate of some assimilations, with marginal active influence of phonotactics.

References

Best, C., Strange, W. 1992. Effects of phonological and phonetic factors on cross-language perception of approximants. *J. Phonetics* 20(3), 305-330. / Dupoux, E., et al. 2011. Where do illusory vowels come from? *J. Memory and Language* 64(3), 199-210. / Flemming, E. 2007. Stop place contrasts before liquids. *Proc. 16th ICPhS Saarbrücken*, 236-240. / Hallé, P., Best, C. 2007. Dental-to-velar perceptual assimilation: A cross-linguistic study of the perception of dental stop+/1/clusters. *J. Acoust. Soc. Am.* 121, 2899–2914.

The usefulness of chaos: Lab versus non-lab speech for perceptual learning

Elizabeth Casserly[1,2] *& David B. Pisoni*[1]

[1]Indiana University, Bloomington IN, [2]Trinity College, Hartford CT

elizabeth.casserly@trincoll.edu, pisoni@indiana.edu

Within the study of speech production, issues of naturalness, ecological validity, and the difference between highly controlled, artificial "lab speech" and everyday production are typically considered (or at least acknowledged) during the processes of data collection and interpretation (e.g. Lindblom, 1990; Xu, 2010). The importance of speech style has made relatively little impact, however, in the complementary study of speech perception. Perceptual research has tended to favor stimulus control over straightforward ecological validity (e.g. Liberman, 1957), although the use of synthesized, artificial speech as stimuli has decreased dramatically since the inception of the field. Currently, many modern perceptual studies make use of recorded utterances from multiple talkers with a variety of backgrounds, in an effort to represent a generalizable range of speech acoustics.

These perceptual stimuli still represent only highly controlled "lab speech," however – and therefore only tell us about how listeners perceive such unusual utterances. Conclusions about how perception is impacted by factors such as inter- and intra-talker variability, noise interference, non-speech environmental sounds, and semantic cohesiveness are therefore weakened, as none of these factors has been studied using stimuli reflecting a listener's typical real-world interaction with these sources of information.

In these studies, therefore, we sought to address this gap in the perceptual literature, comparing and quantifying the effects of unscripted, real-world speech versus typical "lab speech" on naïve listeners' perceptual learning. Specifically, in Experiment 1 (N = 37) we fitted three groups of listeners with a device performing a continuous acoustic simulation of cochlear implant processing, a transformation which substantially degraded the quality and detail of the spectral information for all environmental and speech acoustics. Listeners wore the device throughout the experiment, hearing everything through the degradation. All listeners completed a standard, lab-based test of speech perception accuracy at the beginning of the experiment and again following one hour of structured exposure. The nature of the exposure, critically, varied across the three groups: Subjects in the Lab group received standard automated training, repeatedly responding to clean, context-free recordings of lab speech; subjects in the Active group engaged in one hour of naturalistic conversation with an experimenter; and subjects in the Control group had their signal-transformation device switched off and were asked to transcribe un-degraded lab speech stimuli for an equivalent hour.

Listeners' perceptual accuracy was assessed in the pre- and post-exposure tasks, where they were asked to transcribe non-overlapping sets of 20 pre-recorded IEEE sentences (1969). Transcriptions were scored by the number of designated keywords correctly recognized by subjects. Repeated measures ANOVA on these data revealed significant effects of Group and Exposure, along with a significant Group x Exposure interaction (all p's $< .05$). Post hoc simple tests within each group revealed significant learning effects (pre- vs. post-exposure differences) in the Lab and Active groups (p's $< .05$), with no significant change in Control listeners. Accordingly, corrected pairwise comparisons showed that post-exposure perceptual accuracy was significantly different between the Control and Lab/Active groups (p's $< .05$), but the Lab vs. Active difference in accuracy was not significant ($p > .05$).

Experiment 2 (N = 21) explored areas in which perceptual learning might differ between listeners exposed to lab vs. non-lab speech; the same protocol was used, but a new set of perceptual tasks was administered following the exposure period, including a high-variability speech recognition test and two live-voice isolated word recognition tests. Analysis of the data from Expt. 2 critically first showed a replication of Expt 1: significant and equivalent learning for listeners in the Lab and Active groups relative to Controls (RM-ANOVA, Group x Exp, $p < .05$). Transfer of learning to other perception tasks was not observed, but there was a trend for subjects who heard non-lab conversation to make better use of semantic context in isolated word recognition than subjects exposure to lab speech (univariate post-hoc analyses, Active vs. Lab context benefit score, $p = .044$).

Taken together, the data from Expt. 1 & 2 demonstrate that speech produced outside the control of a phonetics lab is capable of supporting significant perceptual learning. This learning appears to be equivalent to that achieved using tightly-constrained "lab speech," if not enhanced in some circumstances. Given the need for ecological validity in speech perception (as well as speech production) research, therefore, further exploration of the impact of stimuli drawn from real-world production seems to be warranted.

References

IEEE. 1969. IEEE recommended practices for speech quality measurements. *IEEE Trans. Audio Electroacoust.* 17, 227-246. / Liberman, A.M. 1957. Some results of research on speech perception. *J. Acoust. Soc. Am.* 29, 117-123. / Lindblom, B.E.F. (1990). Explaining phonetic variation: a sketch of H&H theory. In: Hardcastle, H.J., Marchal, A. (eds.), *Speech Production and Speech Modeling, NATO ASI Series D: Behavioural and Social Sciences*. Dordrecht: Kluwer A.P., 403-439 / Xu, Y. 2010. In defense of lab speech. *J. Phonetics,* 38, 329-336.

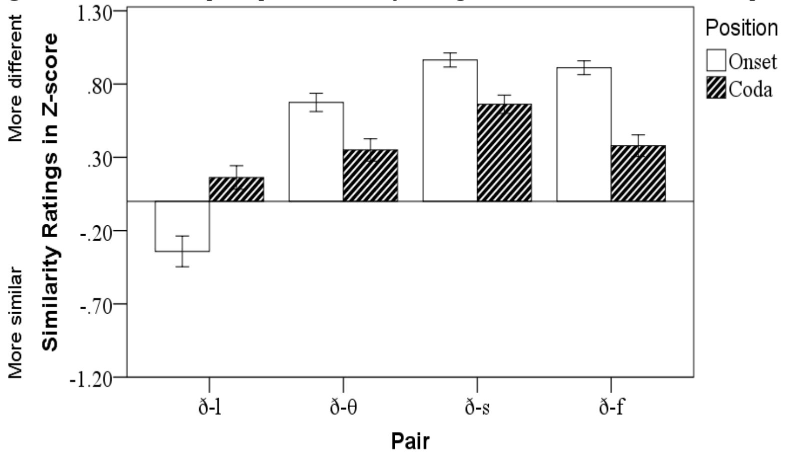
The effect of perceptual similarity in second language learning: Positional asymmetry in phoneme substitution

Yu-an Lu — National Chiao Tung University — yuanlu@nctu.edu.tw
Jiwon Hwang — Stony Brook University — jiwon.hwang@stonybrook.edu

This paper investigates a positional asymmetrical pattern in second language (L2) learning. Taiwanese Mandarin (TM) speakers usually produce the English voiced interdental /ð/ as [l] in onset position (e.g., [ð] *there*→[l]), but as [θ] in coda position (e.g., *loath* [ð]→[θ]) (Lu, 2009). From the first language (L1) transfer point of view, this substitution pattern raises questions as to (i) why voicing is maintained at all, even though voicing is not contrastive in Mandarin (Miao, 2005), (ii) why voicing is maintained only in onset position and (iii) why an illegal coda [ð] is substituted with another illegal coda [θ] when Mandarin does not allow any obstruent coda (Broselow, Chen, & Wang, 1998). Under the universal markedness view, it is still puzzling since more marked [l]/[θ] are preferred to less marked substitutes such as [t]/[n]. This study investigates the effect of *perceptual similarity* to look for a unified explanation for this asymmetrical pattern.

A perceptual similarity experiment was conducted, in which 25 TM speakers (2M, 18F; aged 19-24) were asked to rate how similar [ð] was in comparison with the substitutes [l, θ] and two control sounds [s, f] in onset and in coda positions in three vowel contexts (e.g., [ða-θa], [að-aθ], [lu-su], [ul-us], [si-fi], [is-if]). If TM speakers' substitution pattern of /ð/ is solely driven by perceptual similarity, perceptual similarity between /ð/ and /l/ is expected to be greater in onset than in coda, while that between /ð/ and /θ/ is expected to be greater in coda than in onset. First, the results (Figure 1) showed that the [ð-l] pair was rated as more similar in *onset* than in *coda* whereas the [ð-θ] pair more similar in *coda* than in *onset*, confirming that [ð-l] is perceptually more similar in onset than in coda but [ð-θ] is more similar in coda than in onset. Second, other substitutes [s, f] were generally rated as less similar to [ð] in both positions than [l] and [θ] were, excluding the possibilities of these sounds as potential substitutions. However, two interesting patterns arose from the results. First, [l] was rated as more similar to [ð] than [θ] was, not only in onset position but also in coda position. This is puzzling as to why [l] does not uniformly substitute /ð/ in production because [l] may be a better repair for /ð/ in both onset and coda positions than [θ] is, in terms of perceptual similarity. Second, although [f] was not reported as a possible substitute for /ð/ in Lu (2009), perceptual similarity for the [ð-f] pair was comparable to that for the [ð-θ] pair in coda, suggesting that [f] could be as good as [θ] for coda [ð], if compared only based on perceptual similarity. We propose that perceptual similarity plays a greater role in onset, a perceptually salient position, than in coda, a perceptually weak position, where the L1 grammar emerges. Therefore, in onset, perceptually the most similar phoneme [l] replaces /ð/, while in coda, [θ], which has the same manner feature as [ð] (conforming to the native ranking IDENT-MANNER >> IDENT-VOICE), substitutes it. Similarly, although [f] and [θ] are equally good candidates for coda /ð/ on the perceptual grounds, [θ] is a better substitute for /ð/ on the grammatical grounds because [f] is grammatically more distant from [ð] than [θ] is, in terms of the place feature. The results of this study suggest that the asymmetrical substitution pattern can be fully explained by the interaction of two factors, perceptual similarity and L1 phonology.

Figure 1: Normalized perceptual similarity rating results for [ð-l, ð-θ, ð-s, ð-f] pairs.



References

Broselow, Ellen, Chen, Su-I, & Wang, Chilin. (1998). The emergence of the unmarked in second language phonology. *Studies in Second Language Acquisition, 20*, 261-280. / Lu, Yu-An. (2009). Adaptation of English interdental fricatives by speakers of Taiwan Mandarin. *Current Issues in Unity and Diversity of Languages*, 1151-1165. / Miao, Ruiqui. (2005). *Loanword Adaptation in Mandarin Chinese: Perceptual, Phonological and Sociolinguistic Factors.* Ph.D dissertation, Stony Brook University, Stony Brook.

Sensitivity to fine acoustic detail affects comprehension of reduced speech in L2

Ellen Aalders[1] *& Mirjam Ernestus*[1, 2]
[1]Radboud University Nijmegen, [2]Max Planck Institute Nijmegen
e.aalders@let.ru.nl, Mirjam.Ernestus@mpi.nl

In conversational speech, words are often produced with missing sounds or syllables (e.g., Dalby, 1984). For example, a word like *support* can be pronounced with a shortened or deleted schwa, so that it sounds similar to *sport*. Native (L1) listeners understand such reduced pronunciations with relative ease when they hear them in their semantic context (Ernestus & Warner, 2011). In addition, English listeners can use fine acoustic traces as cues to recover reduced sounds (Manuel, 1991). For second language (L2) learners, understanding reduced pronunciations is more challenging. In the current study we investigate if this is because they are less able to make use of fine acoustic detail in the speech signal.

We focused on schwa-reduction in English and studied how word perception is influenced by listeners' sensitivity to two phonetic cues - schwa duration and aspiration. In English, /p, t, k/ in onsets of stressed syllables are aspirated, but only if they occur in initial position (e.g., in *support* but not in *sport*). Manuel (1991) has shown that English natives use this knowledge during perception. In addition, English native speakers of course attend to the schwa itself.

We compared a group of native English listeners to two learner groups. We first compared them to advanced Dutch learners of English, because Dutch has schwa-reduction (like English) but no aspirated plosives. Secondly, we compared them to advanced Spanish learners of English, who do not know schwa-reduction from L1 and whose native pattern of aspiration also differs (Hualde, 2005). We hypothesized that neither learner group would use aspiration as a cue for word comprehension, and that only the Dutch learners would be sensitive to schwa duration.

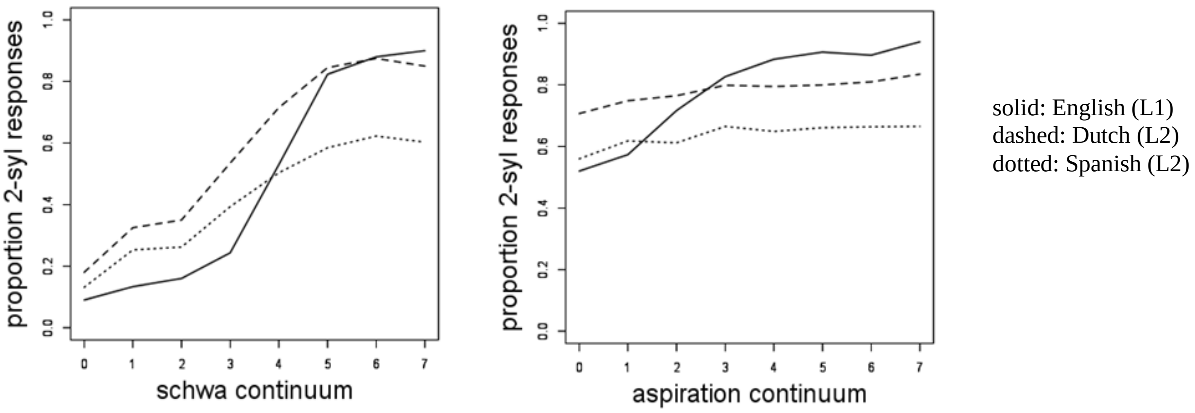
In a decision experiment, participants listened to nonwords and indicated if these were monosyllabic or not. Two continua were made with stimuli ranging from monosyllabic (e.g., *spol*) to bisyllabic (e.g., *suppol*) tokens. In the first continuum we manipulated schwa duration (range: 0 to 48ms); in the second, duration of [p]-aspiration (range: 0 to 65ms). Stimuli were presented in sentence context as well as in isolation.

Logistic regression analyses showed significant differences between the three groups of listeners (see for example Figures 1 and 2). While all groups showed a larger proportion of two-syllable responses with longer schwa durations, the Spanish learners identified words as bisyllabic less often than English listeners, even at long schwa durations. This was true for stimuli presented in context and for stimuli presented in isolation. Furthermore, in sentence context the English natives differed from the Dutch learners, who in turn differed from the Spanish learners. Spanish learners thus have more trouble detecting short schwas than L1 listeners and Dutch learners.

Aspiration of [p] had larger effect on the native speakers than on the two groups of language learners. While all groups interpreted tokens as bisyllabic more often if they contained more aspiration, native listeners were able to use this cue more effectively to recover a reduced schwa.

These results show that problems with the comprehension of reduced speech in L2 can indeed be caused by non-native sensitivity to fine acoustic detail. Exactly how sensitive learners are to fine phonetic cues in their L2 depends on the relevance of these cues in their L1.

Figure 1: Response proportions. The left panel shows the effect of schwa duration for stimuli presented in context. The right panel shows the effect of aspiration duration for stimuli presented in isolation.



References

Dalby, J.M. (1984). Phonetic structure of fast speech in American English. PhD dissertation Indiana University / Ernestus, M., & Warner, N. (2011). An introduction to reduced pronunciation variants. *Journal of Phonetics, 39*(3), 253-60. / Hualde, J. I. (2005). *The sounds of Spanish.* Cambridge: Cambridge UP. / Manuel, S.Y. (1991). Recovery of "Deleted" Schwa. In *PERILUS XIV*, O.Engstrand & C. Kylander (Eds.), 115-118.

Voicing, F0, and phonological enhancement

James Kirby & D. Robert Ladd

University of Edinburgh

j.kirby@ed.ac.uk, bob.ladd@ed.ac.uk

In English, fundamental frequency at vowel onset ('onset F0') is correlated with phonological voicing in initial obstruents: onset F0 following voiceless stops is higher than after voiced stops, and stays higher for longer than would be predicted by physiological or aerodynamic factors alone. It has further been observed that the effect of English 'voiced' (phonetically voiceless unaspirated) stops on onset F0 is the same as that of 'voiced' (phonetically prevoiced) stops in 'true voicing' languages such as French (Hombert, 1978). These findings (among others) led Kingston and Diehl (1994) (KD) to propose that onset F0 differences stem from a gesture intended to produce low F0 as an *enhancement* of [+voice] in stops, regardless of language-particular details of phonetic implementation. However, the idea of a single universal [voice] feature is controversial (e.g. Beckman et al. 2013), and it has long been clear that onset F0 effects depend to some extent on intonational context (Silverman 1986). Hanson (2009), carefully controlling intonational context, showed that onset F0 after English voiced stops is not lower than that of (similarly voiced) nasals, and suggested instead that the effect must be related to *voicelessness*, not voicing. Our study was designed to shed further light on the nature of onset F0 effects by comparing English to languages with prevoiced stops.

Experiment 1: KD's proposal leads us to expect a trading relation between onset F0 and VOT, whereby reduced distinctiveness along the VOT dimension correlates with greater enhancement of onset F0. The implementation of this enhancement should therefore depend on the phonetic detail of [voice]: in 'aspirating' languages like English, longer voicing lag should condition lower onset F0, while in 'true voicing' languages like French, longer voicing *lead* should correlate with *higher* onset F0. We looked for this VOT/F0 covariance in a reading experiment with speakers of English, Italian and French. In English, onset F0 did not change with VOT for [+voice] stops, but (as expected) long-lag VOT in [-voice] stops was accompanied by lower onset F0 (as measured by the VOT:voicing interaction term in a mixed-effects regression; β=-0.44, SE=0.25, $p<0.05$). In Italian and French, however, the opposite effect obtained: there was no relation between VOT and F0 in [-voice] stops ($p = 0.84$), and longer voicing lead was accompanied by *lower* onset F0 ($\beta = 0.08$, SE = 0.03, $p < 0.001$).

Experiment 2: The absence of a trading relation in French and Italian suggests not a phonological but a biomechanical explanation, namely larynx lowering for pharyngeal expansion to maintain phonation. This in turn predicts that [+voice] (prevoiced) stops in French and Italian should condition *lower* onset F0 than sonorants (unlike what Hanson 2009 found for English, where onset F0 is the same after [+voice] (unaspirated) stops and nasals). To test this hypothesis, we replicated a portion of Hanson's (2009) study in Italian (a similar replication with French is underway). We measured F0 in stressed vowels after sonorant /m/, [+voice] /b/ and [-voice] /p/ in two different intonational contexts. Contrary to our expectations, a mixed-effects regression predicting onset F0 from segment type and intonational context found significantly *higher* F0 after /p/ in both contexts (β=23.68, SE=3.79, t=6.25, p<0.0001), and found no F0 difference between /m/ and /b/ (p=0.85) – exactly what Hanson found in English.

These findings support Hanson's idea (2009) that onset F0 is *raised* after [-voice] stops rather than *lowered* after [+voice]. This superficially contradicts KD's proposal that onset F0 enhances the feature [voice], and is difficult to reconcile with the idea of a fundamental phonological distinction between 'aspirating' and 'true voicing' languages. Importantly, however, our results appear consistent with KD's contention that onset F0 effects are driven by phonological rather than purely biomechanical considerations, because they show that segments which are apparently phonetically identical in different languages (viz., 'voiceless unaspirated' stops) induce opposite onset F0 effects depending on their phonological status. Incorporating this finding into any purely biomechanical account of obstruent F0 effects would require us to identify hitherto unsuspected phonetic differences between stops currently classified simply as 'voiceless unaspirated', and/or to find phonetic commonalities in the implementation of voicing contrasts of English, Italian and French.

References

Beckman, J., Jessen, M., Ringen, C. 2013. Empirical evidence for laryngeal features: aspirating vs. true voicing languages. *J. Ling.* 49, 259-284. / Hanson, H. 2009. Effects of obstruent consonants on fundamental frequency at vowel onset in English. *J. Acoust. Soc. Am.* 125, 425-441. / Hombert, J.-M. 1978. Consonant types, vowel quality, and tone. In: Fromkin, V. (ed.), *Tone: a linguistic survey.* San Diego: Academic Press, 77-111. / Kingston, J., Diehl, R. 1994. Phonetic knowledge. *Language* 70, 419-454. / Silverman, K. 1986. F0 segmental cues depend on intonation: the case of the rise after voiced stops. *Phonetica*, 43, 776-791.

[DAY 1: Poster Session 1]

Comparing the discourse use of prosody in adolescents with normal hearing and cochlear implants

Colleen Holt[1], Ivan Yuen[2], David Rosson[2], and Katherine Demuth[2]
University of Melbourne[1] and Macquarie University[2]
cholt@unimelb.edu.au, ivan.yuen@mq.edu.au, david@rosson.com.au, katherine.demuth@mq.edu.au

Previous literature suggests that children fitted with cochlear implants (CIs) have trouble producing appropriate intonation contours. Peng, Tomblin, Spencer & Hurtig (2007) identified the rising tunes used by children with CIs as atypical. They recorded information requests (IR) produced by pediatric CI users in an imitation task, with the model request having a rise of around 12 semitones. Half of the IRs produced by the CI users were judged as having 'no rise', and only 26% were judged as having a 'full rise', with the rise of these productions being around 8 semitones (st).
Given that children fitted with CIs are also reported to have poor discourse/pragmatic skills (Toe, Beattie, & Barr, 2007), we wanted to know whether the use of discourse prosody by this group was also atypical. Previous research using the cooperative Map Task has shown that normal hearing (NH) Australian English-speaking adults routinely use rising tunes in both IR and directives (DIR) (Fletcher, Stirling, Mushin, & Wales, 2002), with higher rises in IR and lower rises in DIR. The goal of the present study was therefore to use the Map Task to determine if prelingually deaf adolescent CI users produce rising tunes in both IR and DIR contexts, and whether these tunes differ quantitatively in excursion or duration from those produced by their NH peers.
The data were drawn from six 12–17-year-olds (Mean 14.5 years), three with unilateral CIs and three with NHs matched for age range. The CI users were prelingually deaf, implanted at 1.3, 3.4 and 4.6 years, with no useable hearing in the other ear, averaging 10.3 years of CI experience. Participants were seated in a sound-treated booth to engage in the Map Task. They were invited to give directions so that the researcher could draw the same route on her map. All productions were recorded at 44.1kHz using a Sony DCM-D50 digital recorder and ECM-T145 condensor lapel microphone. Coding was carried out using Praat. We annotated the beginning and end of speaker turns, rising contours, and the discourse function of each utterance (*directive, information request, acknowledgement, comment*). Rising contours were annotated for rise onset, turning point, and rise offset. The beginning and the end of the word where the rising contour occurred was marked, as well as the syllable preceding it. Given previous studies, we predicted that the participants with CIs might use fewer rises, and/or that these might not be as high.

Overall, the NH speakers took more turns to complete the task (n=33) than the CI users (n=26). Both groups frequently made more than one discourse move within a turn, e.g. issuing a DIR and then following up with an IR, with the NH speakers making more discourse moves than the CI users (39 vs. 30). A chi-square test found a significant difference in the frequency of the discourse moves of interest – DIR and IR (CI vs. NH) χ^2 (1, N = 146) = 9.283, $p < 0.002$). The NH participants made more IR moves than the CI users (31% vs. 15%), whereas the CI participants produced more DIR moves than the NH participants (75% vs. 53%).

The NH and CI speakers produced a comparable proportion of turn-final *rising contours* (59% vs. 56%), with the latter exhibiting large individual differences. However, the two groups also patterned differently as to *when* they used rises. The NH speakers produced rises on both IRs and DIRS (78%, 81%), whereas the CI users produced fewer rises on IRs compared to DIRs (64%, 85%). Phonetically, the realisation of the rises did not differ between the groups; the mean rise excursion for NH IRs was 62 Hz, compared to 69 Hz for the CI users (a difference of 1.9 st). The mean rise of NH DIRs was 77 Hz, and of CI users, 90 Hz (a difference of 2.7 st). Independent *t*-tests conducted on the excursion magnitude of f0 rise and duration of DIRs and IRs separately, with 'groups' as a factor, found no significant differences.

Contrary to our predictions and results of previous studies, the CI users did not differ from the NH speakers in the phonetic realisation of rising contours. This may be due to a combination of early implantation and substantial device experience for these teens. The magnitude of the rises produced by both groups also differed from those reported in Fletcher et al. (2002) in not displaying a difference between DIR and IR rises, suggesting that a follow-up phonological analysis of tunes would be useful to further explore this issue. The main difference between the NH and CI groups in this study was therefore in their approach to the task; the NH speakers used information requests more often, exhibiting more interactive behaviour, whereas the CI users preferred to issue directives and were less interactive with their interlocutor. Future study would therefore want to examine CI users' *perceptual* sensitivity to these discourse uses of prosody, as well as how and when these skills begin to develop.

References

Fletcher, J., Stirling, L., Mushin, I., Wales, R. (2002). Intonational rises and dialog acts in the Australian English map task. *Language and Speech*, 45, 229-253. / Peng, S.C., Tomblin, J.B., Spencer, L.J. & Hurtig, R.R. (2007). Imitative production of rising speech intonation in pediatric cochlear implant recipients. *Journal of Speech, Language and Hearing Research,* 50, 1210-1227. / Toe, D., Beattie, R., & Barr, M. (2007). The development of pragmatic skills in children who are severely and profoundly deaf. *Deafness and Education International*, 9, 101-117.

Code Switch Production in Cantonese-English Bilingual Children

Donald White and Peggy Mok
The Chinese University of Hong Kong
donaldtimothywhite@gmail.com, peggymok@cuhk.edu.hk

In recent years, the phonetics of code switch (CS) production has become a touchstone for examining language interaction in bilinguals. Thus far, research in this area has focused largely on adults (Olson, 2013; Antoniou, 2011; Bullock and Toribio, 2009). While no studies to date have investigated the same phenomenon in children, CS patterns in the early stages of acquisition could provide valuable insights into incipient cross-linguistic patterns manifested by young bilinguals. To this end, the present study is a corpus-based case investigation of CS production in bilingual children (ranging in age from 2;0 to 4;6) acquiring Cantonese and English simultaneously. There are several phonological contrasts between Cantonese and English: Cantonese is a syllable-timed, tone language with relatively simple syllable structure while English is a stress-timed language with more complex phonotactic constraints. In the present study, Cantonese CS tokens found in the Hong Kong Bilingual Child Language Corpus (Yip and Matthews, 2007) of the Child Language Data Exchange System (CHILDES) (MacWhinney, 2000) were compared with the same, or phonetically similar, words spoken by the same children in a monolingual English (ME) context. If language interaction were occurring in the CS tokens, the present study would expect to find intonational, rhythmic, and segmental differences between the CS tokens and their ME counterparts.

Preliminary data from one balanced bilingual child suggest the presence of a base-language effect from Cantonese to English. In the five CS words (a total of 45 tokens) examined, the phonological influence of Cantonese was evident in a number of areas. First, there were several significant differences between the CS tokens and their ME counterparts at the suprasegmental level: the CS tokens were assigned Cantonese tones distinct from the intonational patterns of the ME tokens; syllable juncture was more prominent in CS tokens; and unstressed ME syllables were less reduced when spoken in a CS context. A variety of base-language effects were also manifested at the segmental level. For example, syllable-final dark /l/s were vocalized in CS context as mid-back vowels with significantly higher F1 values; segments in coda clusters were deleted systematically in a CS context; and voiced obstruents in an ME context were often devoiced in a CS context. These findings are currently being corroborated with further data from the other bilingual children.

Collectively, the results suggest both separation and integration between languages of the bilingual children. The nature of this interaction, and its phonological consequences, are dependent on the assumptions regarding the bilingual lexicon. If the bilingual mind contains language-specific lexical entries, then the results suggest that this specificity is separated from the phonological patterning and phonetic implementation of these entries. In this scenario, the CS tokens would seem to be re-encoded phonologically before they are inserted in a base-language utterance. For example, each CS syllable would receive a tone specification bearing no lexical relationship with the words they comprise. If, on the other hand, comprehensive integration is assumed at the lexical and phonological levels, then the results suggest that phonetic implementation is a unique process among bilinguals: lexical items could cross between languages freely, and their phonetic implementation rules would vary according to the language of choice. In this case, the application of tones to CS tokens would represent one possible realization in a phonetic continuum that spans between the two languages.

Whatever their provenance, the patterns observed in the present study have an unambiguous corollary: the cross-linguistic patterns of CS behaviour are entrenched very early in bilingual development. The tendency to adjust CS tokens phonetically seems to be a natural concomitant of early, regular exposure to two distinct phonological systems.

References

Antoniou, M., Best, C., Tyler, M., Kroos, C. 2011. Inter-language interference in VOT production by L2-dominant bilinguals: Asymmetries in phonetic code-switching. *Journal of Phonetics*, 39, 558-570. / Bullock, B., Toribio, A.J. 2009. Trying to hit a moving target: On the sociophonetics of code-switching. In: L. Isurin, D. Winford, & K. De Bot (eds.), *Multidisciplinary approaches to code-switching* [Studies in Bilingualism 41]. Amsterdam: John Benjamins, 189-206. / MacWhinney, B. 2000. *The CHILDES Project: Tools for analyzing talk. Third Edition.* Mahwah, NJ: Lawrence Erlbaum Associates. / Olson, D. 2013. The phonetics of insertional code-switching: suprasegmental analysis and a case for hyper-articulation. *Linguistic Approaches to Bilingualism*, 2(4), 439-457. / Yip, V., Matthews, S. 2007. The bilingual child: Early development and language contact. New York: Cambridge University Press.

Modeling the acquisition of vowel normalization as cognitive manifold alignment

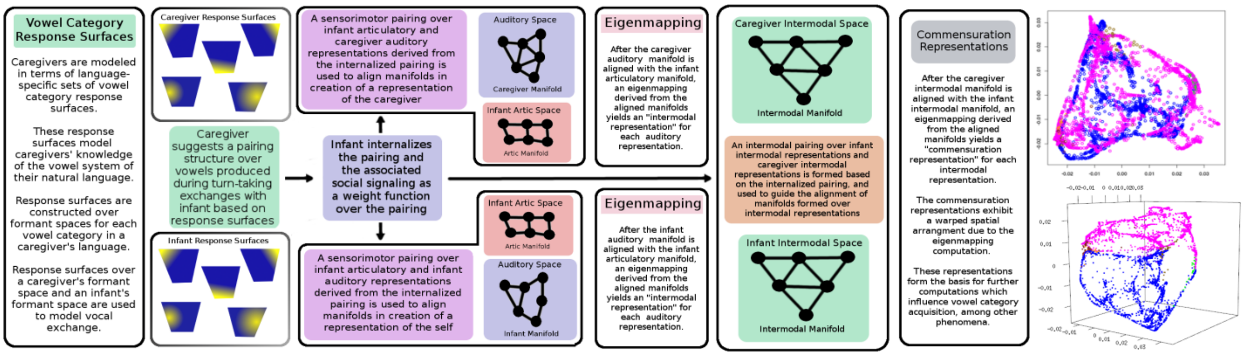
Andrew R. Plummer
The Ohio State University
plummer@ling.ohio-state.edu

Research over the last few decades suggests a reversal of the standard reductive conceptualization of vowel normalization in favor of one which is generative in nature. In this connection, we adopt a cognitive approach to vowel normalization motivated by Gallese's (2001) “shared manifold” hypothesis for the role of mirror neurons in social cognition, which takes a “non-predicative, and non-inferential simulation mechanism, by means of which the observer can recognize and understand the behaviour of others” (p. 44) to be a key aspect in cognitive development. Specifically, we take vowel normalization to be an acquired computation involving infants' formation of models of the self and others, and the formation of relations between these models, based in part on vocal interaction with caregivers.

Manifolds are topological structures which potentially aid in the organization of perception (Seung and Lee, 2000) and speech production (Saltzman, Kubo, and Tsao, 2006), and we take infants' models of the self and others to be cognitive manifolds formed over auditory and articulatory representations, as well as derived “intermodal” and other higher-order representations. We take infants' formation of relations between manifolds to be a cognitive alignment computation which combines two (or more) manifolds into a new aligned manifold based on internalized vocal interaction with a caregiver. Aligned manifolds provide the basis for computing higher-order representations, which may reflect multisensory narrowing (Lewkowicz and Ghazanfar, 2009) and serve as input to other computations, e. g., vowel category acquisition. Within our architecture (depicted in Figure 1), cognitive manifolds are computationally modeled as weighted graphs and cognitive alignment as sets of weighted edges connecting two cognitive manifolds.

We were interested in identifying structural and representational modeling aspects that reflect the influence of language-specific vocal interaction on the acquisition of vowel normalization. We constructed vowel category response surface (Plummer, Ménard, Munson, and Beckman, 2013) models of Greek and Japanese-speaking caregivers based on perceptual categorization experiments (Munson, Ménard, Beckman, Edwards, and Chung, 2010) involving Greek and Japanese-speaking listeners, and corresponding models of infants and their internalization of vocal interaction with their caregivers. Commensuration representations for a Japanese-speaking infant-caregiver dyad are shown at right in Figure 1. The second and third components (top) of these representations exhibit canonical phonological organization, with vowel height along the abscissa, and backness along the ordinate. Inclusion of the first component (bottom) shows the manner in which infant (blue) and caregiver (magenta) representations are pulled together as a consequence of vocal interaction. Examination of the commensuration representations of the remaining Japanese- and Greek-speaking dyads suggests that the acquisition of vowel normalization is not only language-specific, but dyad-specific as well.

Figure 1: Modeling architecture for the acquisition of vowel normalization. Acoustic and articulatory representations are from an articulatory synthesizer (Boë, Perrier, and Girin, 2010). Auditory representations are ERB transformations (Moore, Glasberg, and Baer, 1997) of vowel signals. (right) Commensuration representations for a Japanese-speaking dyad.



References

Boë, L.-J., Perrier, P., and Girin, B. 2010. Vlab Software package. Université du Québec à Montréal / Gallese, V. 2001. The ‘shared manifold’ hypothesis. From mirror neurons to empathy. *J. Conscious. Stud.*, 8(5-7):33-50. / Lewkowicz, D. J. and Ghazanfar, A. A. 2009. The emergence of multisensory systems through perceptual narrowing. *Trends Cogn. Sci.*, 13:470-478. / Moore, B. C. J., Glasberg, B. G., and Baer, T. 1997. A model for the prediction of thresholds, loudness, and partial loudness. *Journal of the Audio Engineering Society*, 45(4):224–240. / Munson, B., Ménard, L., Beckman, M. E., Edwards, J., and Chung, H. 2010. Sensorimotor maps and vowel development in English, Greek, and Korean: A cross-linguistic perceptual categorizaton study (A). *J. Acoust. Soc. Am.*, 127:2018. / Plummer, A. R., Ménard, L., Munson, B., and Beckman, M. E. 2013. Comparing vowel category response surfaces over age-varying maximal vowel spaces within and across language communities. *Proc. of INTERSPEECH*. / Saltzman, E., Kubo, M., and Tsao, C. C. 2006. Controlled variables, the uncontrolled manifold, and the task-dynamic model of speech production. In: Divenyi, P., Greenberg, S., and Meyer, G., (eds.), *Dynamics of Speech Production and Perception*, 21-31. / Seung, H. S. and Lee, D. D. 2000. The manifold ways of perception. *Science*, 290(5500):2268–2269.

Individual differences in children's prosodic focus-marking

Aoju Chen
Utrecht University & Max Planck Institute for Psycholinguistics
aoju.chen@uu.nl

Background: Past work has shown that the use of prosody to mark focus (i.e. new information in a sentence) is acquired in stages at the group level with phonological cues being acquired before phonetic cues [1]. Individual differences in development have however been observed among children [1]. The important question of what causes such differences has received very limited attention. Are individual differences in the ability to use prosody to mark focus purely a matter of phonological and phonetic competence? Or are they related to development in other areas?

Aim: This study aims to address these questions by examining whether variation in children's prosodic focus-marking in production can be accounted for by variation in children's comprehension of the prosody-to-focus mapping, verbal intelligence and musicality. Recent work suggests that children's ability to use prosody to mark focus goes in tandem with their ability to process the prosody-to-focus mapping [2]. This suggests that children better in comprehension may also be better in production. Verbal intelligence is the ability to analyse information and solve problems using language-based reasoning. High verbal intelligence indicates high general intelligence; low verbal intelligence is associated with poor language development [4]. A recent study on Chinese-learning 4- to 5-year-olds showed that verbal intelligence was positively correlated with accuracy in prosodic focus-marking [3]. It remains to be seen whether this is applicable to Dutch children. Further, better musical abilities appear to predict better pitch-related prosodic abilities in language production and perception [5]. Others have however argued for lack of connection between musical and prosodic abilities [6]. It is thus not clear whether musicality is related to the ability to use prosody in focus-marking.

Method: Seventy-five typically-developing Dutch 4- and 5-year-olds (age range: 4;1 to 5;11; mean age: 5;3) participated in this research. In the production experiment, we elicited from the children SVO sentences (as responses to who/what-questions) with focus on subject-NPs (initial focus), and object-NPs (final focus) using a picture-matching game adopted from [1]. The usable full-sentence responses and corresponding questions were subsequently combined into context-response dialogues. Three intonationally-trained native speakers of Dutch listened to the dialogues and evaluated each response on how well its prosody fitted in the context on a 5-point scale ("1": does not fit; "5": fits perfectly). Second, the children participated in a comprehension experiment [2], in which the children judged answers of some animals to who/what-questions raised by a boy about a number of pictures. The question-answer dialogues here were syntactically comparable to the dialogues in the production experiment; the answer sentences were of similar length across focus conditions. The reaction times (RT) to sentences with appropriate prosodic focus-marking and sentences with inappropriate prosodic focus-marking were recorded. The children's verbal intelligence was assessed via the Peabody Picture Vocabulary Test for Dutch (PPVT-III-NL). Their musical abilities were assessed via the Primary Measures of Music Audiation (PMMA) test, frequently used in research on the development of musical abilities.

Analysis & Results: Verbal intelligence scores and musicality scores were obtained for each child following the standard procedures. In addition, a 'production' score was computed for each child in each focus condition by averaging the scores of available responses in each focus condition. The production scores ranged from 1 to 5, indicating substantial individual differences. A 'comprehension' score was computed for each child by calculating the ratio between the mean log-transformed RT in the inappropriate-prosody condition and that in the appropriate-prosody condition in each focus condition. If a child responded faster in the appropriate-prosody condition than in the inappropriate-prosody condition, he would have a comprehension score bigger than 1. Linear regression analysis was conducted on the data from the children who had all four kinds of scores. The models revealed that only musicality was a significant predictor, accounting for 22% of the variation in production in the final focus condition ($r=.466$, $p = .033$).

Conclusions: Our results have shown that individual differences in production were not related to variation in verbal intelligence, different from [3], suggesting that verbal intelligence may have language-specific effect on acquisition of prosodic focus-marking. Its relevance to Chinese children's prosodic focus-marking points to a possible connection between a larger vocabulary and better control of post-lexical use of pitch. Further, individual differences in production were not related to comprehension, suggesting independence in producing prosody and comprehending prosody, contra our expectation. Finally, musical abilities did play a role in accounting for the individual differences in prosodic production, providing evidence for the connection between musical and prosodic abilities. However, considering that musical abilities could only account for one fifth of the variation in production, we suggest that the acquisition of prosodic focus-marking may be primarily a matter of developing phonological and phonetic competence.

References: [1] Anonymous. Tuning information packaging: Intonational realization of topic and focus in child Dutch. *J. of Child Language*. [2] Anonymous. Is there really an asymmetry in the acquisition of the focus-to-accentuation mapping? *Lingua*. [3] Anonymous. Individual differences in children's focus marking: verbal intelligence and theory of mind. Paper presented at TIE5. [4] Groth-Marnat, G. (2009) *Handbook of psychological assessment* (5th ed.). New York: Wiley. [5] Patel, A.D. & Iversen, J.R. (2007) The linguistic benefits of musical abilities. *Trends in Cognitive Sciences*, 11, 369-372.[6] Peretz, I. (2009) Music, language and modularity framed in action. *Psychological Belgica*, 49(2-3): 157-175.

Effects of speaker language and listener language on children’s stop place

Mary E. Beckman, Benjamin Munson, Jan Edwards

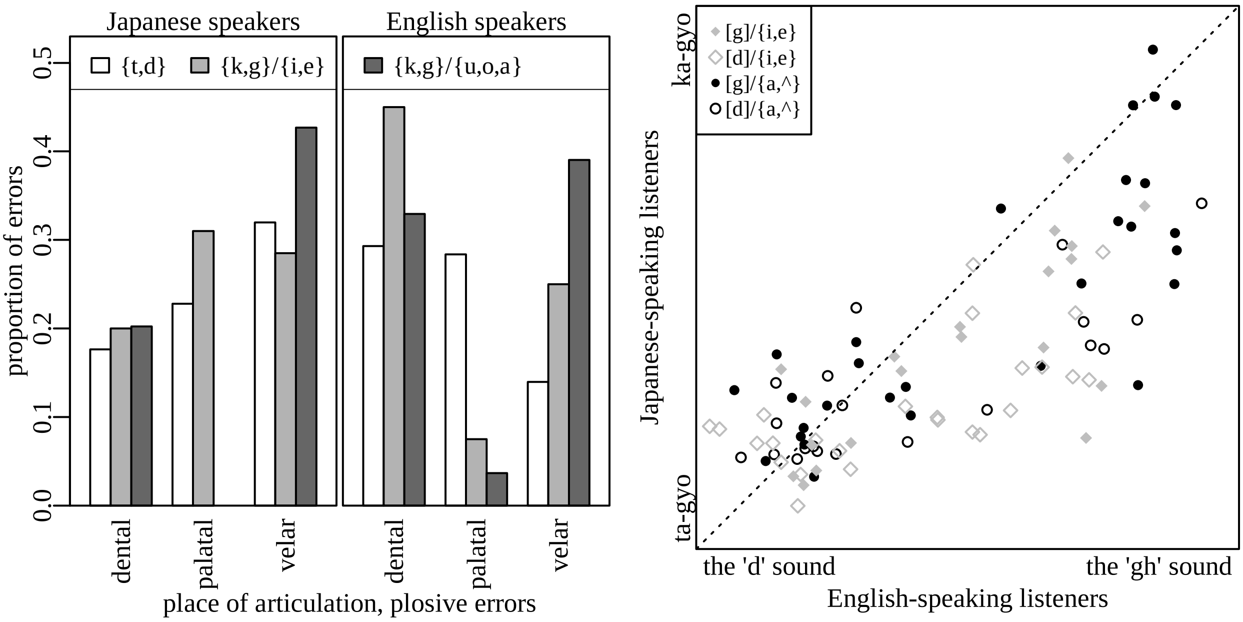
Ohio State University, Columbus; University of Minnesota, Twin Cities; University of Wisconsin, Madison

mbeckman@ling.osu.edu, munso005@umn.edu, jedwards2@wisc.edu

At least as early as Jespersen (1922), linguists have described *velar fronting* – the perceived substitution of a more anterior place of articulation for target [k] or [g] – in young children who speak languages such as Danish, German, and English. As the left panel of Figure 1 shows, in English-speaking children, velar fronting tends to be a substitution of [t] or [d] rather than of a palatal affricate for the target velar, and it is observed in both front and back vowel environments. The same was true of the children observed by Jesperson, and the prevalence of this stereotypical error is sometimes cited as evidence that dental/alveolar place of articulation is universally “unmarked” relative to velar place. In Japanese-speaking children, however, fronting errors for velars are more typical of front vowel contexts, where they tend to be transcribed as substitutions of [tʃ] or [dʒ] rather than of [t] or [d]. Moreover, a more stereotypical error for Japanese-speaking children is backing of [t] and [d] to [k] or [g] in back vowel environments, an error that is rarely seen in English-speaking children. Beckman, Yoneyama, and Edwards (2003) suggest that these differences between the stereotypical error patterns for the two languages might be due to the different relative frequencies of the stop types in different vowel contexts, and the generally more anterior articulatory setting for English relative to Japanese.

If this suggestion is correct, we might predict that Japanese- and English-speaking adults differ in how they perceive young children’s productions of stops. We tested this prediction by presenting 23 English- and 20 Japanese- speaking adults with CV stimuli extracted from children’s productions of words such as *tisshuu* ‘tissue’ and *kame* ‘turtle’ (for the Japanese stimuli) and *deer* and *garden* (for the English stimuli). The stimuli included both correct and incorrect productions, with the incorrect productions for each language selected to represent the stereotypical errors of both languages – e.g., taking most of the alveolar-backing errors of English and a sample of the same number of velar-fronting errors. The listeners rated each stimulus on a visual analog scale that was anchored by “the ‘d’ sound” vs. “the ‘gh’ sound” for the English speakers and by a vowel-appropriate kana symbolization (e.g., “た” vs. “か” for *kame* and *garden*, “ティ” vs. “き” for *tisshuu*) for the Japanese. The right-hand right panel of Figure 1 shows the results for all of the English stimuli, plotting the average rating by the Japanese-speaking listeners against the average rating by the English-speaking listeners. Productions of correct [d] and of fronted [g] were rated similarly, but productions of correct [g] and of backed [d] were consistently rated as more front (closer to the dental anchor) by Japanese-speaking listeners.

Figure 2: Distribution of transcribed error types for word-initial [t, d] vs. [k, g] in front vowel contexts vs. [k, g] in back vowel contexts in a corpus of word productions elicited from Japanese- and English-speaking children (left) and place ratings by Japanese- and English-speaking listeners for a subset of the English-speaking children’s productions with correct or incorrect place (right).



References

Beckman, M.E., Yoneyama, K., Edwards, J. 2003. Language-specific and language-universal aspects of lingual obstruent productions in Japanese-acquiring children. *J. Phon. Soc. Japan* 7, 18-28. / Jespersen, O. 1922. *Language: Its Nature, Development, and Origin*. New York: Henry Holt and Company.

The acquisition of differential spectral kinematics of English sibilant fricatives

Patrick F. Reidy

The Ohio State University, Dept. of Linguistics

reidy@ling.ohio-state.edu

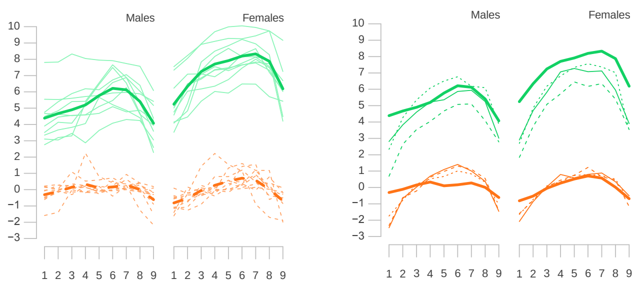
Background. Iskarous *et al.* (2011) investigated the spectral kinematics of English /s/ by approximating the trajectory of the first spectral moment (centroid) across each token's duration. Growth-curve models fit to these trajectories were found to have a significant positive linear coefficient and a significant negative quadratic coefficient, indicating that the centroid follows a convex downward, increasing trajectory. Despite its dynamic spectral properties, static measures of the dominant frequency, such as centroid or peak frequency (e.g., Jongman *et al.*, 2000), computed from a single location within the fricative, have traditionally proved sufficient for distinguishing /s/ from the other English sibilant /ʃ/. These static measures have furthermore been used to show that the degree to which /s/ and /ʃ/ are differentiated, varies across fluent adult speakers (Newman *et al.*, 2001) and age and gender demographics (Romeo *et al.*, 2013).

Purpose of study. Representing /s/ and /ʃ/ categories in terms of one or more static spectral measures admits a valid description of the /s/-/ʃ/ contrast only in so far as those spectral measures follow similar trajectories across the duration of both sibilants. Thus, evidence of significant curvature in the centroid trajectory of /s/ poses the questions of whether /ʃ/ exhibits similar spectral kinematic patterns, and, if not, of the developmental trajectory of children as they acquire the ability to produce the appropriate adult-like spectral patterns for each sibilant.

Methods. Productions of /s, ʃ/ in word-initial, pre-vocalic position were elicited from native English-speaking adults ($N = 20$) and native English-acquiring three-, four-, and five-year-old children (N = 14, 18, 18, respectively) with a picture-prompted, real word-repetition task. Trained phoneticians transcribed each token phonemically, and only phonemically correct tokens were included in the analysis. The method for analyzing each token's spectral kinematics was similar to that of Iskarous *et al.* (2011): Nine 20-ms rectangular windows were spaced evenly throughout the duration of each token, and from each window, the spectrum of the fricative was estimated with an 8^{th}-order multitaper spectrum (Thomson, 1982). However, rather than reducing each spectral estimate to its centroid (as was done by Iskarous *et al.*, 2011), a measure of the peak psychoacoustic frequency was computed by passing each spectral estimate through a gammatone filterbank model of the auditory system (Patterson, 1976), which output a psychoacoustic spectrum-like object (similar to the "excitation pattern" of Moore *et al.*, 1997), and then finding in this output the psychoacoustic frequency with maximum amplitude, which is referred to as the *peakERB* (Holliday *et al.*, 2011).

Results. Adults did not produce the same spectral kinematic pattern for /s/ and /ʃ/: the peakERB of /s/ followed a convex downward, increasing trajectory, while for /ʃ/ it remained relatively flat (Figure 1a). Furthermore, the children exhibited a developmental path toward the peakERB trajectory appropriate to each sibilant: the three-year-olds produced /s/ and /ʃ/ with comparable curvature in their peakERB trajectories; however, as age increased, the children produced /ʃ/ with decreasing curvature to its peakERB trajectory, approaching that produced by the adults (Figure 1b).

Figure 3: Normalized peakERB (y-axis) trajectories across analysis windows (x-axis). *Left panel:* Individual (thin, light lines) and demographic (thick, dark lines) mean peakERB trajectories for the adults' productions of /s/ (green, solid lines) and /ʃ/ (orange, dashed lines). *Right panel:* Mean peakERB trajectories for three-, four-, and five-year-olds (thin solid, dashed, dotted lines, respectively), relative to the adult trajectories (thick solid line).



References

Holliday, J., Beckman, M.E., Mays, C. 2010. Did you say susi or sushi? Measuring the emergence of robust fricative contrasts in English- and Japanese-acquiring children. *Proc. 11th InterSpeech*, 1886-1889. / Iskarous, K., Shadle, C.H., Proctor, M.I. 2011. Articulatory-acoustic kinematics: The production of American English /s/. *JASA.* 129(2), 944-954. / Jongman, A., Wayland, R., Wong, S. 2000. Acoustic characteristics of English fricatives. *JASA.* 108(3), 1252-1263. / Moore, B.C.J., Glasberg, B.R., Baer, T. 1997. A model for the prediction of thresholds, loudness, and partial loudness. *J. Aud. Eng. Soc.* 45(4), 224-240. / Newman, R.S., Clouse, S.A., Burnham, J.L. 2001. The perceptual consequences of within-talker variability in fricative production. *JASA*. 109(3), 1181-1196. / Patterson, R.D. 1976. Auditory filter shapes derived with noise stimuli. *JASA*. 59(3), 640-654. / Thomson, D.J. 1982. Spectrum estimation and harmonic analysis. *Proc. IEEE*. 70, 1055-1096.

Perception and production of phonemic vowel length in Australian English-learning 18-month-olds

Hui Chen, Nan Xu Rattanasone and Felicity Cox
Department of Linguistics, ARC Centre of Excellence in Cognition and its Disorders, Macquarie University
hui.chen15@students.mq.edu.au, nan.xu@mq.edu.au, felicity.cox@mq.edu.au

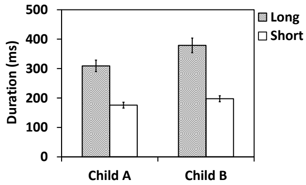
Vowel quantity serves important functions in the acquisition of phonological/prosodic systems in many languages. It can be used to distinguish word meanings (e.g., *cart* /ɐː/ vs. *cut* /ɐ/ in Australian English (AusE)), and can contribute to syllable weight, thereby influencing the stress patterns of words.

Vowel quantity is suggested to be mastered later than vowel quality (e.g., Fikkert, 1994; Demuth & Fee, 1995; Buder & Stoel-Gammon, 2002). For example, Japanese infants are reported to discriminate vowel quality differences earlier than durational differences (Sato, Sogabe, & Mazuka 2010). However, there are few studies of children's early awareness of vowel quantity at the word learning stage (Dietrich, Swingley, & Werker, 2007), and little agreement can be reached on whether vowel quantity contrasts are produced before the age of two (Fikkert, 1994; Ota, 1999; Kehoe & Lleó, 2003) since it is difficult to get enough comparable data from young children. Studies have shown that by 18 months, children can identify a single feature change in vowel quality (Mani, Coleman, & Plunkett, 2008), but it remains unclear whether 18-month-olds can also *identify* and *produce* contrastive vowel quantity.

This study therefore addresses the question of whether infants can distinguish vowel *quantity* distinctions in both perception and production by 18 months. We selected AusE as the target language since it has two contrastive vowel length pairs - eː/e and ɐː/ɐ, which presumably makes the acquisition process even harder compared to the mastery of vowel quantity in the languages that have systematic long/short contrasts. We used the Intermodal Preferential Looking (IPL) paradigm to test sensitivity to the mispronunciations of vowel quantity and vowel quality in AusE-acquiring 18-month-olds with Tobii Eye Tracker X120. Vowels /e, æ, ɐ, ɜː, ɐː/ were tested in 20 CVC words with a voiceless coda. Six children (3M; 3F) were presented two pictures of different objects side by side, one familiar and one novel, and heard an auditory prompt asking them to look at one of the objects. The auditory prompt could be a correct label or a mispronunciation of the familiar word, or a label completely different from the familiar word. The results show that, by 18 months, AusE-learning children are aware of the distinctions in both vowel length and vowel height in a word recognition task.

In the production test, we engaged the same children in an elicited imitation task. Participants were asked to repeat the minimal pairs of the CVC words with a voiceless stop coda that were used in the perception test, and these were recorded using Audacity and Behringer C20 microphone at 48K Hz. Sufficient data were collected from two of the children (1M; 1F) (/ɜː, ɐː/: 11 tokens from Child A and 7 from Child B; /e, ɐ/: 12 from Child A and 6 from Child B. We used Praat to measure vowel durations from the onset and offset of F2 in the spectrogram. One-tail paired t-tests show that both children produced significant differences between target long and short vowels (Child A: $t(10)=1.813$, $p<.001$; Child B: $t(5)=2.015$, $p=.002$). The results reveal that, by 18 months, at least those toddlers who were willing to talk can produce phonemically contrastive vowel length in AusE. This suggests that vowel *quantity* contrasts may be achieved earlier in production than often thought, even in phonological systems where this contrast is not found systematically throughout the vowel inventory, as it is in Swedish, Finnish, Arabic or Japanese.

Figure 4: Mean durations (ms) of target long vs. short vowels produced by two AusE 18-month-olds.



Reference

Buder, E.H., & Stoel-Gammon, C. 2002. American and Swedish children's acquisition of vowel duration: effects of vowel identity and final stop voicing. *J. Acoust. Soc. Am.*, 111 (4), 1854-1864. / Demuth, K., & Fee, E.J. 1995. *Minimal words in early phonological development*. Ms., Brown University and Dalhousie University. / Dietrich, C., Swingley, D., & Werker, J. 2007. Native language governs interpretation of salient speech sound differences at 18 months. *PNAS*, 104 (41), 16027-16031. / Fikkert, P. 1994. *On the acquisition of prosodic structure*. PhD dissertation, University of Leiden, published by Holland Institute of Generative Linguistics. / Kehoe, M.M., & Lleó, C. 2003. The acquisition of nuclei: a longitudinal analysis of phonological vowel length in three German-speaking children. *J. Child Lang.*, 30, 527-556. / Mani, N., Coleman, J., & Plunkett, J. 2008. Phonological specificity of vowel contrasts at 18-months. *Lang. and Speech*, 51 (1&2), 3-21. / Ota, M. 1999. *Phonological theory and the acquisition of prosodic structure: Evidence from child Japanese.* PhD dissertation, Georgetown University. / Sato, Y., Sogave, Y., & Mazuka, R. 2010. Discrimination of phonemic vowel length by Japanese infants. *Dev. Psychol.*, 46, 106-199.

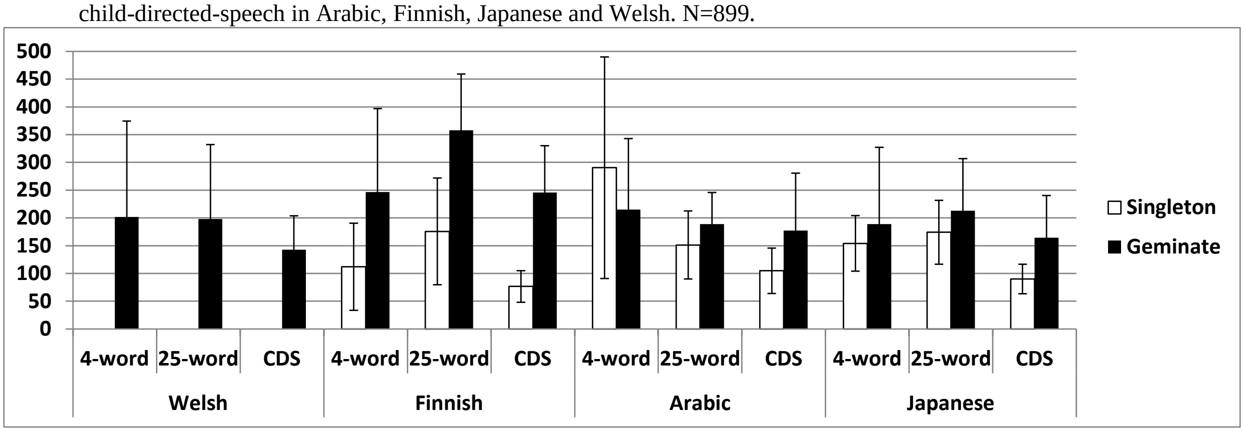
A comparison of long consonant acquisition in Arabic, Finnish, Japanese and Welsh

Ghada Khattab, Marilyn Vihman**, Jalal Al-Tamimi*, Satsuki Nakai^, Sari Kunnari^^*
*Newcastle University, **University of York, ^Glasgow University and ^^University of Oulu
ghada.khattab@newcastle.ac.uk

The acquisition of contrastive consonant length is an understudied area in phonology despite the window it offers into the child's attunement to the prosody of her language and the transition from phonetic ability to phonological knowledge. One of the challenges involved in the production of target-like consonant duration in the early stages of word production is the child's slow moving articulators, which typically lead them to produce phonetically long consonants regardless of the language they are acquiring (Vihman et al 2006); this is normally followed by a process of shortening as their articulatory control matures. The initial production of phonetically long consonants is at odds with the fact that that geminates are cross-linguistically rare, leading to predictions of an unmarked (i.e. short C) starting point for the child. Instead, the challenge lies in contrastive use of length, which requires both phonetic control (i.e. to produce and perceive both short and long targets in different prosodic contexts) and phonological representation (to express meaning, morphological patterns, etc). While early acquisition of consonant length can be predicted in languages where gemination has a heavy functional load, other factors may play a role in providing a stable model for the child. These include the degree of variability in consonant realization as a function of linguistic features such as stress, syllable structure, the frequency of occurrence of geminates in the language and the ratio of singleton to geminate consonants; they also include external factors such as the degree of variability in the input to children.

In this study we explore factors affecting the acquisition of medial consonants length in three languages with geminates (Arabic, Finnish, and Japanese) and one language with prosodically-governed long consonants (in trochaic disyllables) but no phonological length contrast. Finnish and Welsh offer the most stable context for long consonant acquisition for very different reasons: Finnish disyllables are all trochaic, the ratio of C to CC is 1:3, the majority of consonants can be geminates, and the realization of singleton (S) and geminate (G) targets in the input is stable (Kunnari et al 2001). Welsh trochaic syllables are systematically realized with long consonants, and the iambic patterns with short Cs are rare (Kunnari & Vihman, 2006). Geminates in Arabic have a lexical and morphological role and are frequent in the language, but the ratio of C to CC is close to 1:2 and input to children is extremely variable (Khattab & Al-Tamimi, 2013). As for Japanese, while gemination has a lexical role it is more often used for pragmatic purposes (Aoyama, 2000), the ratio of C to CC is similar to Arabic, and input is equally variable. Spontaneous data from half-hour monthly recordings of 5 children from each language between the beginning and end of the one-word period (referred to here as 4wp and 25wp) demonstrates phonetic and phonological effects on the acquisition of long consonants in these languages (Fig.1). While all children initially produce long consonants across the board, an emerging S-G contrast for Finnish children is already present at the 4wp, and their stops are generally longer. Welsh children show the most stability across the two developmental points, reflecting the more stable target in the adult language. Arabic and especially Japanese children's productions exhibit more variability and are still short of the adult target at the 25wp.

Figure 5: Mean and SD values of medial stop duration (in ms) in disyllabic words at the 4-word and 25-word stage in children and in child-directed-speech in Arabic, Finnish, Japanese and Welsh. N=899.



References

Aoyama, K. 2000. Acquiring mora-timing: the case of the Japanese coda nasal. *Jpn-Korean Ling*, *9*. /Khattab, G.& Al-Tamimi, J. 2013. Early phonological patterns in Lebanese Arabic. In Vihman, M. M., & Keren-Portnoy, T. (Eds.). *The Emergence of Phonology* Cambridge: CUP. / Kunnari, S., Nakai, S., & Vihman, M. M. 2001. Cross-linguistic evidence for the acquisition of geminates. *Psychology of Language and Communication*, *5*(2), 13-24. / Vihman, M. M., & Kunnari, S. 2006. The sources of phonological knowledge: A cross-linguistic perspective. *Recherches linguistiques de Vincennes*, (35), 133-164. /Vihman, M. M., Nakai, S., & DePaolis, R. A. 2006. Getting the rhythm right: A cross-linguistic study of segmental duration in babbling and first words. *Laboratory phonology*, *8*, 341-366.

Do infants normalize speaker and accent variability in vowel production?

Karen E. Mulak, Samra Alispahic and Paola Escudero
MARCS Institute, University of Western Sydney, Australia
k.mulak@uws.edu.au, s.alispahic@uws.edu.au, paola.escudero@uws.edu.au

Speech produced by different speakers and across different accents contains striking acoustic variation, but adults are able to hear through it with relative ease. Our aim is to establish whether or not the same underlying mechanisms are at play when we cope with speaker versus accent variation. There is evidence suggesting that these processes may be different but most models of speech processing seem to treat them as the same type of *indexical* variation. While speaker *normalization* appears to be a basic automatic process that even non-human animals such as Chinchillas and Zebra finches perform (e.g., Kuhl & Miller, 1975; Ohms et al. 2010), resolving accent variation typically requires exposure, suggesting it may be a learned linguistic process (e.g., Floccia et al., 2006). Infant speech perception research demonstrates that vowel discrimination across different speakers is in place at 7 months (Kuhl, 1983), but other studies show that infants have difficulty recognizing words if spoken by a different speaker (Houston & Jusczyck, 2000).

The current study aims at examining whether infants cope with speaker and accent variation equally by presenting 7-month-old Australia English (AusE) infants with vowels produced by speakers from two different accents of a foreign language. We chose two Dutch vowels whose productions largely differ in Dutch spoken in North Holland versus Dutch spoken in East Flanders, namely /ɪ/ and /ɛ/. AusE infants are unlikely to be familiar with any of the speakers of these accents but will likely be able to distinguish the vowels since they also exist as different phonemes in English. If speaker variation is more negligible than accent variation, infants should more readily notice a change in one of the vowels if produced by a speaker of a different accent than if produced by a different speaker of the same accent.

Method

Twenty-four 6.6- to 20.3-month-olds (M = 11.4 mos; SD = 5.3 mos; 13 females) from Sydney, Australia were exposed to eight familiarization trials each with 16 repetitions of the vowel /ɪ/ produced by a female speaker of North Holland Dutch. Infants were randomly assigned to two testing conditions were they heard three test trials: 1) another familiarization trial (Same), 2) the familiarized vowel /ɪ/ alternating with the vowel /ɛ/ spoken by the same female speaker (NewVowel), and either 3a) the familiarized /ɪ/ alternating with /ɪ/ spoken by another female with the same accent (NewSpeaker, N = 13), or 3b) the familiarized /ɪ/ tokens alternating with /ɪ/ tokens produced by a female of East Flemish Dutch (NewAccent, N = 11). Participants' gaze to a colorful bull's eye that displayed on the screen when vowel stimuli played was recorded using a Tobii X120 eyetracker sampling at 120 Hz. E-Prime 2.0 was used to present participants with the eight familiarization trials and two blocks of three randomly presented test trials (Condition 1: Same, NewVowel, NewSpeaker; Condition 2: Same, NewVowel, NewAccent) in a serial preference procedure. All trials began with an attention video, with each trial starting once participants were fixated on the screen.

Results

Difference scores for test trials were calculated by subtracting the looking to the screen during the Same test trials from their looking time to each of the other three test trials. For each condition, a repeated-measures ANOVA on test trial difference (NewVowel-Same = 2.00 sec in Condition 1 and 0.30 sec in Condition 2, NewSpeaker-Same = 1.90 sec or NewAccent-Same = 3.20 sec) was conducted and revealed that the difference compared to the Same trials (i.e., test trial equal to those presented during familiarization) was greater for NewAccent trials compared to NewVowel trials, $F(1, 10) = 6.22$, $p = .032$. There was no difference between NewSpeaker and Same trials. A further one-sample t-test showed that the difference between NewAccent and Same trials approached significance, $t(10) = 1.77$, $p = .080$, whereas the other differences did not. Further, independent-samples t-test showed that the NewVowel-Same difference was not significant across conditions, $t(22) = .556$, $p = .584$.

Discussion

The aim of this experiment was to see whether changes in accent are handled differently to changes in speakers of the same accent. While participants did not detect the speaker change, the results suggest that they did detect a change in accent. This suggests that speaker and accent variation is not handled in the same manner. Children did not detect a vowel change, which may be due to the Dutch vowel contrast not being the same as that in AusE. Alternatively while this suggests that speaker and accent variation are not handled the same way, it may be related to the specific productions of the speakers used in the present experiment. Ongoing research is testing more infants in different age groups and different variations of the vowel, speaker and accent changes.

References

Floccia C, Goslin J, Girard F, Konopczynski G (2006) *J Exp Psych: HPP, 32*, 1276-93/ Houston D, Jusczyk P (2000) *J Exp Psych: HPP, 26*, 1570–82/ Kuhl P (1983) *Infant Beh & Dev, 6,* 263-285/ Kuhl P, Miller J (1975) *Science, 190*, 69-72/ Ohms V, Gill A, van Heijningen C, Beckers G, ten Cate C (2010) *Proc R Soc B, 277*, 1003-09.

On lexical phonotactics and segmentability

Robert Daland[1], Benjamin Börschinger[2], Abdellah Fourtassi[3]

[1]Department of Linguistics, UCLA; [2]Department of Computing, Macquarie University; [3]Laboratoire de Sciences Cognitives et Psycholinguistique, ENS/EHESS/CNRS, Paris

{r.daland, abdellah.fourtassi}@gmail.com , benjamin.borschinger@mq.au.edu

Word segmentation is the perceptual process(es) by which infant and adult listeners parse speech into a sequence of word-sized units, even in the absence of previous experience with those words. A variety of computational models of word segmentation have been proposed, relying on phonotactics (e.g. Daland & Pierrehumbert, 2011), lexical coding (e.g. Goldwater et al., 2009), or some hybrid approach (e.g Fleck, 2008). Since there are excellent electronic resources for English, most computational models are initially applied to English; dismayingly it has been found that in all cases where a segmentation model is applied to another language, the performance drops relative to English, sometimes drastically so, and regardless of whether the model is more phonotactically or lexically oriented (Daland & Zuraw, 2013; Fleck, 2008; Fourtassi et al., 2013). This fact is dismaying because it is generally assumed that the developmental trajectory of word segmentation does not differ strongly across languages, which implies that the right model of word segmentation should exhibit a similar robustness to cross-linguistic variation; this correspondingly suggests that no extant model is even close to right. However, it is also possible that the modeling results are telling us something important -- that languages really do (or can) vary in their segmentability, as suggested independently by Daland & Zuraw (2013) and Fourtassi et al. (2013). These two papers studied Korean and Japanese, respectively, and proposed that the lower segmentability (relative to English) arose from phonotactic properties, such as the restricted syllable inventory. The present paper explores this line of reasoning further by generating artificial languages in which the phonotactic structure is systematically varied, while controlling other factors that might influence segmentability, such as the word frequency distribution.

Method. Artificial language corpora were generated in three stages: defining a *grammar*, generating a *lexicon/corpus* according to a frequency distribution, and then applying an unsupervised word *segmentation* algorithm. *Grammars*. The alphabet ∑ consists of a 'typical' segmental inventory of 13 consonants [ptkfsxmnŋlɾwj] and 5 vowels [aeiou]. Grammars were stated in the Maximum Entropy Harmonic Grammar formalism, in which constraints are stated as illicit sequences of natural classes using SPE-style featural descriptors, e.g.the constraint banning all consonant clusters is written '*[-syl][-syl]'. The 'harmony' of a string in ∑* is the weighted sum of its constraint violation, and the probability of a string is proportional to the exponential of its harmony (Hayes & Wilson, 2008). The maximally restricted language, **Pseudo-Senufo**, is defined as a strict CV language (words cannot begin with V; words cannot end with C; no CC; no VV). The maximally unrestricted language, **Pseudo-Berber**, includes no phonotactic constraints at all (so that the form [tfktkx] would be just as probable as the form [patiku]). An intermediate language, **Pseudo-Korean**, includes phonotactic properties reminiscent of Korean (maximally CGVC syllables, coda neutralization, [l]/[ɾ] allophony, Syllable Contact Law). *Lexicons/corpora*. Lexicons were sampled from the probability distribution over ∑* assigned by the grammar, subject to a maximum word length of 6 and a modest penalty on every segment (*Struct) which served as a soft preference for shorter words. Lexical types were then assigned frequencies, and tokens were scambled, so as to match the type-token and utterance-length distributions in a subset of the Brent-Bernstein-Ratner corpus. *Segmentation*. The Adaptor Grammar formalism assigns a posterior distribution over segmentations of the corpus, according to a generative model of the language (Johnson et al., 2007). Sample segmentations are generated according to a Markov Chain Monte Carlo process, and the final segmentation is generated by applying Minimal Bayes Risk decoding to posterior samples. The token f-scores are as follows: Pseudo-Berber 93%; Pseudo-Korean 92%; Pseudo-Senufo 83%.

Discussion. The results suggest that when distributional factors are controlled, languages which are more phonotactically restricted are predicted to be harder to segment, consistent with previous proposals (Daland & Zuraw, 2013; Fourtassi et al., 2013). At present, it is unclear whether every phonotactic constraint is always predicted to reduce segmentability; for example, differential rates of assimilation between and across word boundaries might make some phonotactic cues stronger (Daland & Pierrehumbert, 2011). This abstract reports on 3 grammars; the final results are expected to include additional languages, such as a variant of Pseudo-Korean with the Syllable Contact Law constraints removed, and variants of Pseudo-English with stronger and weaker degrees of Sonority Sequencing Principle for onsets.

References. Daland, R. & Pierrehumbert, J.B. 2011. Learning diphone-based segmentation. *Cog. Sci.*, 35(1), 119-155. / Daland, R. & Zuraw, K. 2013. Does Korean defeat phonotactic word segmentation? *Proc. 51st ACL, Sofia, Bulgaria*. / Fleck, M.M. 2008. Lexicalized phonotactic word-segmentation. *Proc. 45th ACL: HLT*, 130-138. / Fourtassi, A., Börschinger, B., Johnson, M., & Dupoux, E. 2013. Whyisenglishsoeasytosegment? *Proc. 51st ACL, Sofia, Bulgaria*. / Goldwater, S., Griffiths, T.L. & Johnson, M. 2009. A Bayesian framework for word segmentation: Exploring the effects of context. *Cognition* 112, 21-54. / Johnson, M., Griffiths, T.L., & Goldwater, S. 2007. Adaptor Grammars: A framework for specifying compositional nonparametric Bayesian models. *NIPS* 19: MIT.

Learning sound categories with phonotactics

Masaki Noguchi and Carla Hudson Kam
University of British Columbia
mskngch@gmail.com, Carla.HudsonKam@ubc.ca

A great deal of recent experimental work investigates the role of distributional learning in the acquisition of phonology, focusing mostly on sound categories and phonotactics. In these studies, however, phonotactics and sound categories are treated as separate phenomena despite the fact that we know that they are highly interdependent. Phonotactics partly determines whether two sounds are separate phonemes or allophones; they are phonemic when their occurrences are unpredictable from their phonological contexts, while they are allophonic when their occurrences are predictable from their phonological contexts (Hall 2009). In this study we investigate the role of context in the distributional learning of sound categories. In particular, we examine the role of sequential regularities in speech, or phonotactics, in sound category learning through an artificial language learning experiment modelled on those conducted by Maye and colleagues (Maye 2000, Maye et al. 2002)

Studies on speech perception have suggested that allophones are perceptually less distinguishable than phonemes (Boomershine et al. 2008), and we made use of this finding in our study. We tested whether the presentation of the input data such that the occurrences of target sounds were predictable from their phonological contexts (i.e., the target sounds were in complementarily distribution and so were like allophones) leads to those sounds being less distinguishable as compared to when they were presented such that they were not predictable from their phonological contexts.

Method: 40 adult native English speakers were tested on their learning of two novel fricative categories, retroflex [ʂ] and alveolo-palatal [ɕ]. The experiment consisted of two sessions spread over two consecutive days. In each session participants heard about 15 minutes of input in which there was a bimodal distribution of stimuli drawn from an 8-step continuum of fricative sounds. The end points of the continuum, step 1 and step 8, were the canonical tokens of retroflex [ʂ] and alveolo-palatal [ɕ] respectively, but the frequency peaks of the distribution were located at step 2 and step 7. Crucially, half of the participants heard stimuli in which the occurrences of the target categories were predictable; retroflex [ʂ] occurred after [u] and alveolo-palatal [ɕ] occurred after [i] (Complementary (C) condition) while the other half heard stimuli in which the occurrences of the target categories were unpredictable; both target sounds occurred before [u] and [i] (Non-complementary (NC) condition). In both sessions, participants took an AX discrimination test post-exposure, in which they compared stimuli from two points along the continuum that should be assigned to different categories if participants had extracted phoneme-like categories. Participants were tested with the comparison of end point tokens, step 1 and step 8 (distant pair) and the comparison of acoustically more similar tokens, step 3 and step 6 (close pair).

Results: Figure 1 shows the results of discrimination test in d-prime scores. (1) participants showed a better sensitivity to distant pair than close pair. (2) participants showed an improvement in their sensitivity from session 1 to session 2. (3) participants in the NC condition showed higher sensitivity to the difference between target categories than those in the C condition. The last point was observed across test stimuli pairs and sessions. This clearly suggests that the difference in phonotactic context or the predictability of target categories in input affects whether learners treat the target sounds as being in the same or different categories.

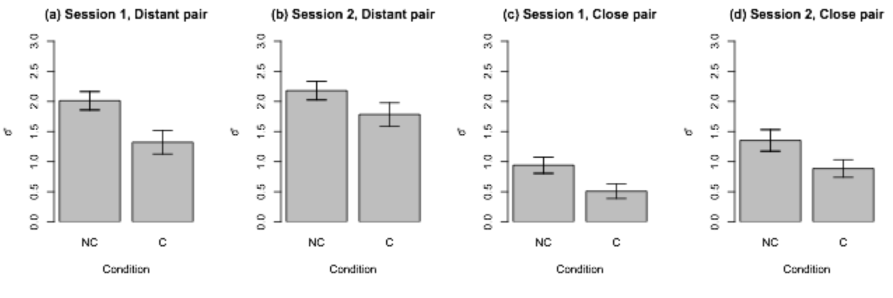


Figure 1: d’ scores for AX discrimination test (NC=Non-complementary group, C=Complementary group)

References

Boomershine, A., Hall, K. C., Hume, E., and Johnson, K. (2008). The influence of allophony vs. contrast on perception: the case of Spanish and English. In P. Avery, B. E. Dresher, and K. Rice (Eds.), Contrast in phonology: perception and acquisition./ Hall, K. C. (2009). A probabilistic model of phonological relationships from contrast to allophony. Ph.D. dissertation. The Ohio State University./ Maye, J. (2000).The acquisition of speech sound categories on the basis of distributional information. Ph.D. dissertation. The University of Arizona./ Maye, J., Werker, J. F., and Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. Cognition, 82 (3), B101-B111.

What kinds of units are used in speech perception?

Eva Reinisch
Dept. of Phonetics and Speech Processing, Ludwig Maximilian University Munich, Germany
evarei@phonetik.uni-muenchen.de

Listeners use lexical or visual context information to recalibrate auditory speech perception. After hearing an ambiguous sound between /f/ and /s/ in lexical contexts where it can only be interpreted as /f/ but not /s/ (e.g., at the end of "giraffe"), listeners recalibrate their phonetic categories such that in lexically ambiguous contexts (e.g., knife-nice) the ambiguous sound is interpreted in line with the previously experienced lexical context. Similarly, after hearing an auditory stimulus between "aba" and "ada" coupled with a clear visual stimulus (e.g., lip closure in "aba"), an ambiguous auditory-only stimulus is perceived in line with the previously seen visual stimulus. Phonetic recalibration has been shown with various phoneme contrasts and has recently been proposed as a means to study the units underlying speech perception (e.g., Mitterer, Scharenborg, & McQueen, 2013) by testing generalization of learning.

Candidate units suggested for recalibration are abstract phonemes (due to generalization of recalibration across positions in the word and across words) or allophones (due to lack of generalization across allophonic pronunciation variants). However, Kraljic and Samuel (2006) showed that listeners generalize recalibration of duration cues to voicing from alveolar stops (/d/-/t/) to labial stops (/b/-/p/). They suggest that listeners recalibrate phoneme-independent durational information rather than the specific phoneme contrast. Alternatively, listeners could have recalibrated perception of abstract gestures involved in stop voicing. The present study set out to further test this suggestion for phoneme-independent recalibration of acoustic cues or gestures by exploring generalization across phoneme contrasts differing in manner of articulation. That is, recalibration of /p/-/t/ or /b/-/d/ and generalization to /m/-/n/ was assessed.

The labial-alveolar contrast is partly cued by the formant trajectories of the second and third formants in adjacent vowels. These cues are similar for stop and nasal consonants. Following the paradigm used by Kraljic and Samuel, a lexically-guided phonetic recalibration experiment was first conducted. Eleven /p/ and /t/-final German words were selected and presented in a lexical decision task for exposure (note that German has final consonant devoicing). One group of listeners heard stimuli in which /p/ was replaced with an ambiguous sound between /p/ and /t/ and another group heard stimuli in which /t/ was replaced with an ambiguous sounds. Ambiguous stimuli were created via audio morphing using STRAIGHT. Following this exposure listeners performed a phonetic categorization task categorizing sounds along a 6-step [a:p]-[a:t] as well as an [a:m]-[a:n] nonword continuum (in that order). Listeners showed the expected recalibration for the stop contrast (between-group difference /p/-responses 16%) but no generalization to the nasal contrast (difference 0.7%). It thus seems that phonetic recalibration of a place contrast is dependent on manner of articulation. The mere labial vs. alveolar cues or gestures appear not to be the unit of recalibration.
However, in this lexically-guided recalibration experiment ambiguous stimuli were created by audio morphing, that is, the simple overlay of the two signals. Therefore a second experiment using visually-guided recalibration was run in which the acoustic cues to the critical contrasts could be tightly controlled. Note that the workings of lexically-guided and visually-guided recalibration appear to be similar (van Linden & Vroomen, 2007). Following the methods of previous visually-guided recalibration experiments, listeners were presented with short exposure-test blocks. During exposure listeners saw eight video clips of a speaker producing "aba" or "ada" paired with an ambiguous auditory stimulus. Immediately thereafter they categorized an audio-only continuum containing the three most ambiguous steps of an original "aba"-"ada" continuum to test the basic recalibration effect or an "ama"-"ana" continuum to test generalization. Note that the cues to place of articulation were manipulated to be identical across the stop and nasal continua. That is, the direction of formant trajectories of F2 and F3 in the vowels surrounding the critical sounds was the only informative cue. The stop portion of the signal was set to silence and the nasal portion of the signal was set to an ambiguous value between /m/ and /n/. Importantly, listeners clearly recalibrated their perception of the /b/-/d/ continuum (difference 11%), but despite identical cues to place of articulation they did not generalize this recalibration to the nasal contrast (difference 1.5%). Similarly, if listeners were exposed to videos of the speaker articulating either /m/ or /n/ paired with an ambiguous nasal, listeners recalibrated the nasal contrast (difference 19%) but did not generalize recalibration to the stop contrast (difference <0.1%). The lack of generalization was thus not due to the fact that nasal contrasts cannot be recalibrated.

In summary, listeners recalibrate a place of articulation contrast by means of lexical or visual information but they do not generalize this recalibration across manner of articulation - even if the cues are identical between the two contrasts. The units of speech perception affected in phonetic recalibration thus appear more specific that abstract acoustic cues (i.e., direction of formant trajectories) or abstract gestures. Note that if abstract gestures were the units of speech perception, effects of generalization would have been expected at least in the visually-guided recalibration paradigm. Place of articulation was then not only cued by identical acoustic cues but also the visual gesture of lip closure.

References

Kraljic, T., Samuel, A. 2006. Generalization in perceptual learning for speech. *Psychon. Bull. Rev,* 13, 262-268 / Mitterer, H., Scharenborg, O., McQueen, J. M. 2013. Phonological abstraction without phonemes in speech perception. *Cognition,* 129, 356-361. / Van Linden, S., Vroomen, J. 2007. Recalibration of phonetic categories by lipread versus lexical information. *J. Exp. Psychol. Human,* 33, 1483-1494.

Phonological and phonetic vowel reduction in two Ibero-Romance varieties

Marianna Nadeu

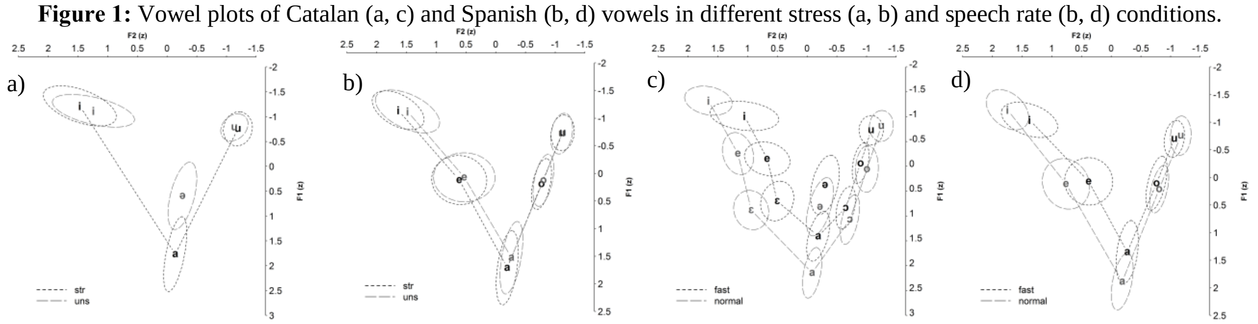
Penn State

nadeu@psu.edu

It is common for lexical-stress languages to show stress-induced vowel quality variation. Some languages exhibit a less crowded vowel system in unstressed position due to certain vowel alternations and neutralizations (phonological vowel reduction). In other languages, stress causes only slight phonetic variation, typically vowel centralization and shortening (phonetic vowel reduction). In fact, de Jong's (1995) *hyperarticulation hypothesis* (HH) predicts that linguistically encoded information (e.g. vowel place features) will be emphasized in prosodically prominent positions. This study explores how prosodic factors (lexical stress, pitch accent, speaking rate) affect vowel quality and duration in Central Catalan (Cat) and Iberian Spanish (Sp), which, albeit closely related, present important phonological differences. Due to phonological vowel reduction, the 7 stressed vowels of Cat (full vowels: /i, e, ɛ, a, ɔ, o, u/) alternate with only 3 in unstressed syllables (reduced vowels: [i, ə, u]). Sp, on the other hand, has 5 vowels (/i, e, a, o, u/) in both positions.

Study 1: 20 speakers of Cat and 20 speakers of Sp read non-words in which the target vowel appeared in the first syllable surrounded by /p/. The words were disyllabic, and the position of stress was varied to elicit stressed and unstressed tokens. In task 1, participants read the carrier sentences at their normal speech rate and, in task 2, at a faster rate. Each speaker produced 100 vowels (10 vowels * 2 rates * 5 repetitions). A total of 1996 vowels for Sp and 1987 for Cat were entered in the analysis (repeated-measures ANOVAs and t-tests with Bonferroni correction to explore the interactions). Fig. 1 shows the effects of stress on vowel quality in Cat (a) and Sp (b) and those of speaking rate in Cat (c) and Sp (d). Statistical analysis showed that all Cat vowels shorten and centralize in F2 (and some also in F1) in unstressed position (in line with HH) and at faster speech rate. In Sp, faster speaking rate also results in shortening and centralization. Yet, the lack of a consistent pattern across speakers indicates that stress does not have a conventionalized effect on vowel quality in Sp. Unstressed vowels are shorter like in Cat, but the effect is smaller and less stable in Sp.

Figure 1: Vowel plots of Catalan (a, c) and Spanish (b, d) vowels in different stress (a, b) and speech rate (b, d) conditions.



Study 2: In study 1, the effects of stress in Cat could only be examined for /i, u/. Full vowels in unstressed position occur in exceptional contexts, e.g. in verb + noun compounds ([ˈkasə] '(it) catches' + [ˈmoskəs] 'flies' = [kasəˈmoskəs] 'fly-catcher'), which preserve the stress of the second root only (Mascaró, 1983; Prieto, 2003). Study 2 investigated whether these exceptional unstressed full vowels differ in duration and quality from lexically stressed vowels and from unstressed, reduced ones by comparing verb + noun compounds with segmentally homophonic verb phrases (stressed condition) and morphologically related words (unstressed condition) in accented and deaccented contexts. 20 speakers produced 252 vowel tokens each (7 vowels * 3 stress conditions * 2 contexts * 2 accent conditions * 3 repetitions). 4776 vowels were included in the analyses. In both conditions, unstressed full vowels were shorter and formed a more compressed vowel space than stressed full vowels. Absence of a pitch accent on lexically stressed vowels did not affect their quality or duration. Thus, vowel quality and duration are correlates of lexical stress, not pitch accent in Cat.

These studies show that even closely related languages may differ on whether they possess phonological rules of vowel reduction, but also in the conventionalization of specific processes of phonetic vowel reduction. This discovery has implications for our understanding of the nature of phonetic processes and how they relate to phonological rules.

References

de Jong, K. 1995. The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *J. Acoust. Soc. Am.* 97, 491-504. / Mascaró, J. 1983. *La fonologia catalana i el cicle fonològic*. Bellaterra: Publicacions de la Universitat Autònoma. / Prieto, P. 2003. Correlats acústics de l'accent secundari en català. *Estudios de fonética experimental* 12, 106-142.

The effect of English stress on non-native speakers' word recognition: Evidence from native speakers of lexical tone language

Shu-chen Ou
National Sun Yat-sen University, Taiwan
sherryou@mail.nsysu.edu.tw

This study investigates whether tone language speakers use the cue of stress to recognize English words. Previous studies have indicated that the cue of stress is not very useful for native speakers of English to distinguish stress minimal pairs (e.g., hearing *fór-* to distinguish *forbear* (n.) vs. *forbear* (v.)) (Cutler, 1986). However, non-native speakers whose mother tongue is typologically related (e.g., Dutch) do use it to recognize English words (e.g., hearing *mú-* and then deciding whether *music* (or *museum*) is a word or not), and they even outperform native speakers of English (Cooper et al, 2002; Cutler et al. 2007). No study has been done using non-native speakers whose mother tongue is typological distant from English (e.g., Mandarin Chinese, a lexical tone language). This study therefore examines this issue. The cue of pitch varies within a syllable to mark word meaning in Mandarin Chinese, so its speakers are supposed to be sensitive to English stress, especially when stress is realized by high/low pitch differences.

A cross-modal priming experiment was carried out: (i) the sound file of the first syllable of a test word was played (e.g., *cám-* of *campus*), (ii) a visual stimulus was shown in the center of the screen (e.g., *campus*), and then (iii) a decision had to be make about whether the visual stimulus was a real English word or not by pressing by pressing the YES button or the NO button as soon as possible. Twelve pairs of words with stress differences were chosen (e.g., *campus – campaign*, *current - correct*, *river – reveal*, *infant –inform*). Twenty-five L1 Mandarin speakers (experimental group) at high English proficiency level (i.e., TOEIC scores > 850) and 25 native speakers of English (control group) were recruited to participate in the experiment. The main findings are as follows. First, for both groups, the words with initial stress were recognized faster in the match condition (e.g., hearing *cám-* and seeing *cámpus*) than in the mismatch condition (e.g., hearing *cam-* and seeing *cámpus*), suggesting that the cue of strong stress facilitates the recognition of words with initial stress). Second, words with final stress were recognized slightly faster in the match condition than in the mismatch condition in the experimental group, suggesting that the cue of weak stress somehow facilitates the process of word recognition, a result not found in the control group. In summary, both the cues of strong stress and weak stress facilitate word recognition by tone language speaking learners of English, and the effect of strong stress is greater. This study provides evidence that the cue of stress is employed in online word recognition by speakers whose native language is typologically distant from English.

References

Cooper, N., Cutler, A. and Wales, R. (2002). Constraints on Lexical Stress on Lexical Access in English: Evidence from native and non-native listeners. *Language and Speech*, 45, 207-228.

Cutler, A. (1986). Forbear is a homophone: Lexical prosody does not constrain lexical access. *Language and Speech 29*(3), 201-220.

Cutler, A., Wales, R., Cooper, N., & Janssen, J. (2007). Dutch listeners' use of suprasegmental cues to English stress. *Proceedings of the 16th ICPhS* (pp. 1913-1916)

Evidence of early motor planning in Speech Production

Donald Derrick (a,b), Romain Fiasson (a)
(a) New Zealand Institute of Language, Brain, and Behaviour, University of Canterbury
(b) MARCS Institute, University of Western Sydney
donald.derrick@canterbury.ac.nz, romain.fiasson@canterbury.ac.nz

Evidence has been found for speech planning at the level of syllables (Levelt, 1994), phonemes (Levelt, 1989; Dell, 1986) and features (Dell, 1986; Mowrey and MacKay, 1990; Bernhardt and Stemberger, 1998). Researchers (Anonymous, submitted) also demonstrated subphonemic planning spanning morpheme/word boundaries. They demonstrated this subphonemic planning by first identifying, through ultrasound of the tongue tip/blade, four categorically different tongue gestures for English flap/taps: An upward "up-flap" motion ([ɾ↖]), a downward "down-flap" motion ([ɾ↘]), an up-down "alveolar tap" motion ([ɾ↕]), and a front-back "postalveolar tap" motion ([ɾ↔]). For “editor”, the initial flap is more likely to be an up-flap, in anticipation of a tip-up rhotic vowel at the end of the word, than for “edify”, demonstrating tongue motion planning based on expected productions across a morpheme boundary.

This paper builds upon and extends the latter findings by examining motor planning as a function of speech rate. Here we report findings on tongue height as measured on the EMA signal. We asked the following questions: (1) How early in the speech chain can evidence of motor planning be found? (2) Is motor planning dependent on speech rate? (3) If so, to what degree does speech rate influence motor planning?

Three American English speakers were asked to produce 10 blocks of 8 VɾVɾV words/phrases, fully contrasting rhotic and non-rhotic vowels, (e.g. “editor” and “edit a”) at 5 different speech rates. Before each trial, participants heard reiterant speech matched in syllable count to the tokens and played at one of 5 rates from 1.5 to 3.5 feet per second. They were then instructed to read aloud a token sentence at the same speech rate. Synchronized audio, ultrasound, and EMA (of the tongue and lips) data were collected. EMA was head corrected, and audio segmented to isolate the vowels and flaps. EMA data for tongue-tip height was then extracted according to the audio segmentations, and a procrustean fit applied to the data to align the mid-point of the initial vowel, first flap contact as identified by amplitude minimum (Zue, 1979), mid-point of the middle vowel, second flap contact, and mid point of the final vowel. The data was pooled and the tongue tip heights by words compared via SSANOVA. Here we present a comparison of “edit a”, ending in a non-rhotic vowel, and “editor”, ending in a rhotic vowel, as seen in Figure 1.

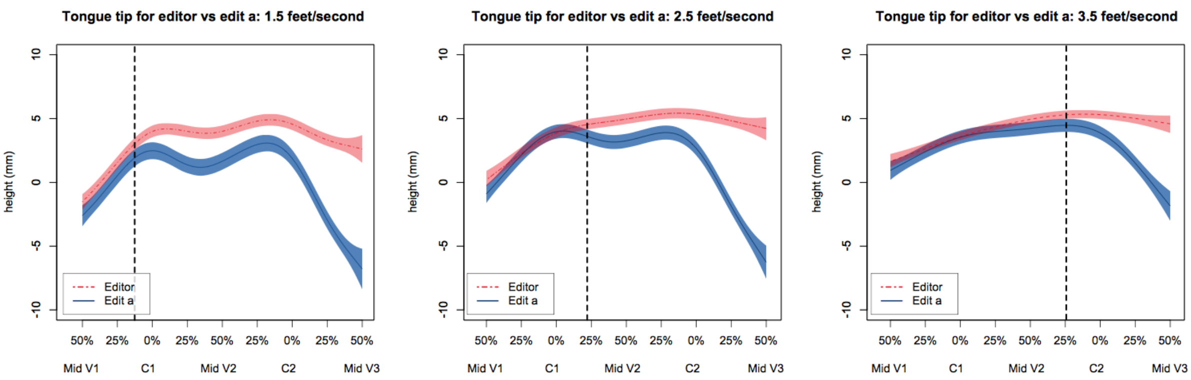


Figure 1: Tongue heights for “editor” vs. “edit a” by speech rate (Vertical dashed line = beginning of sig. divergence)

The results show gradient low-level motion planning of the tongue tip based on upcoming planned speech as early as the first vowel in a VɾVɾV sequence (1). The results are speech rate dependent (2). In slow speech, this occurs as early as the initial vowel in the morpheme “edit” in relation to the final vowel in the word “editor” or the final vowel in the phrase “edit a”, and in faster speech significant divergence occurs just prior to the final flap variant (3). This is evidence for gradient speech motor planning across morpheme and word boundaries, dependent upon speech rate.

References

Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation*. MIT Press, Cambridge, Massachusetts / Levelt, W. J. M. (1994). Do speakers have access to a mental syllabary? *Cognition*, 50:239–269 / Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93(3):283–321/ Mowrey, R. A. and MacKay, I. R. A. (1990). Phonological primitives: Electromyographic speech error evidence. *JASA*, 88(3):1299–1312 / Bernhardt, B. H. and Stemberger, J. P. (1998). *Handbook of phonological development from a nonlinear constraints-basted perspective*. Academic Press, San Diego, CA / Zue (1979), Acoustic study of medial/t,d/in American English, *JASA*, 66(4): 1039-1050

Acoustic effects of predictability and gender on Japanese high vowel reduction

James Whang
New York University
james.whang@nyu.edu

The main purpose of this study is to investigate influences on the amount of cues available in the acoustic signal of reduced high vowels in Japanese. In standard Tokyo Japanese, the reduction of unaccented /i, u/ in a C1-V-C2 sequence when both C1 and C2 are voiceless has received significant interest in the fields of phonetics and phonology. However, there is disagreement regarding how the reduced vowels are manifested acoustically, if at all. Some argue that the reduced vowels are devoiced, retaining oral gestures that overlap with and thus color the burst or frication noise of C1 despite the loss of voicing (Varden, 2010). Others argue that the most robust acoustic cue for reduced vowels is an extra long burst or frication noise of C1 that signal the presence of an underlying vowel (Han, 1994). Still others argue that the vowels simply delete because tokens containing reduced vowels have been lexicalized as such, making such acoustic cues unnecessary (Ogasawara, 2013). The results of the present study show that the predictability of a vowel determines the degree of reduction. Also, apparent lengthening of C1 burst/frication is shown to depend on an extra-linguistic factor, namely gender.

Data used in this study were obtained from eight native Japanese speakers (four women and four men) who were born and raised in the Tokyo area. The participants were recorded in a sound proof booth reading 40 lexical tokens (20 target and 20 control) embedded in meaningful and unique carrier sentences. Both target and control tokens were divided into stop-stop and fricative-stop combinations. For all tokens, C1 was [k, ʃ], after which either high vowel can occur, or [ɸ, s, ç], after which only one of the two is possible. C2 was [k, t, ʃ, ɸ, s, ç] for target tokens and [b, d, g, z] for control tokens. The recorded data were segmented using Praat. Center of gravity (COG), the amplitude weighted mean of frequencies present in a signal, was measured for the first half (COG1) and the second half (COG2) of C1 burst/frication noise to determine the degree of vowel coloration. Greater coarticulation with a vowel would lower the amplitude of the higher frequencies. The duration of C1 burst/frication noise in reduced target tokens and unreduced control tokens were compared to check for any length differences.

COG measurements revealed that the cues available in the acoustic signal depend on the recoverability of the vowel in a given context. For [ɸ, s, ç], after which the vowel is predictable, COG2 was significantly lower than COG1 when an unreduced vowel followed, suggesting an increased overlap towards the end of C1. In contrast, no such effect was found for reduced tokens, suggesting that there is no intervening oral vowel gesture between C1 and C2. In other words, the vowel deleted, leaving no cue for the vowel because any such cues are not necessary for recovering an already predictable vowel. In contrast, for [k, ʃ], after which the vowel is unpredictable, COG1 was significantly lower for both reduced and unreduced tokens when the vowel was /u/. Being a back vowel, /u/ has a larger front oral cavity than /i/ and thus lower resonance frequencies. The significant effect of vowel type suggests that the vowels did not delete but devoiced, where the retained vowel gesture overlaps with and colors the burst/frication noise of C1. Complete deletion of the vowel in these cases would jeopardize the recoverability of the vowel. By devoicing the vowel instead (i.e., reducing the glottal gesture), maximal recoverability is obtained.

Measurements of C1 burst/frication duration showed a significant effect of vowel reduction only for [k], where the stop burst was longer for reduced tokens. None of the fricatives showed such an effect. A possible explanation for the difference between [k] and the fricatives is that the burst of [k] is simply too short to carry enough information for reliable recovery of the reduced vowel, and thus lengthens unlike the fricatives, which are already long enough. This difference, however, was driven entirely by the female participants, with the male participants showing no such lengthening effect. This finding is consistent with Imai (2010), who reports that women produce significantly less vowel reduction than men. Stated differently, acoustic information for an underlying vowel is retained more by women than men. While puzzling at first glance, this gender-based speech difference is not surprising especially in Japanese, where systematic gender-based differences exist at all levels of the language (Inoue, 2002). These findings are consistent with previous research arguing that gender should be considered as a factor in phonetic analyses more broadly (Ryalls, Zipprer, & Baldauff, 1997).

References

Imai, T. 2010. An emerging gender difference in Japanese vowel devoicing. In: Preston, D.R., Niedzielski, N.A. (eds.), *A Reader in Sociophonetics*. Walter de Gruyter, 177–187. / Inoue, M. 2002. Gender, language, and modernity: Toward an effective history of Japanese women's language. *American Ethnologist*. 29, 392–422. / Han, M.S. 1994. Acoustic manifestations of mora timing in Japanese. *J. Acoust. Soc. Am.* 96, 73–82. / Ogasawara, N. 2013. Lexical representation of Japanese high vowel devoicing. *Language and Speech,* 56:5–22. / Ryalls, J., Zipprer, A., and Baldauff, P. 1997. A preliminary investigation of the effects of gender and race on voice onset time. *J. Speech Lang. Hear. Res.* 40, 642–645. / Varden, J.K. 2010. Acoustic correlates of devoiced Japanese vowels: Velar context. *J. Eng. Am. Lit. Ling.* 125, 35–49.

Tapping in American English: Context matters

Adam J. Chong and Megha Sundara
University of California, Los Angeles
ajchong@ucla.edu, megha.sundara@humnet.ucla.edu

How adults recognize variant forms is central to research on spoken word recognition. We are interested in variant forms that are conditioned by the segmental context. It is now well established that adults' recognition of words fails when assimilated variants are presented out of context (e.g. Gaskell & Marslen-Wilson, 1996; Otake, Yoneyama, Cutler & van der Lugt, 1996; Gow, 2003; Mitterer & Blomert, 2003). The role context plays in adults' recognition of words with perceptually distant variant forms is more contentious (for - Pitt, 2009; against - Ranbom *et al*., 2009). In this paper, we present data from 2 experiments to challenge Ranbom *et al*.'s claim that because context does not provide a reliable cue for tapping in word-final context in American English, its recognition proceeds without regard to segmental context. As in Ranbom *et al*., we investigated adult native English speakers' recognition of *t*-final words. Unlike Ranbom *et al*., we did not provide listeners with a semantically biasing sentence context. This was done to ensure that listeners' performance was not at ceiling. Instead, listeners were presented with variant forms in two semantically neutral sentence frames: *Click the word X now!* or *Click the word X again!*

In Experiment 1, American English listeners were tested on a lexical decision task (n=14). They heard target words with stops (e.g. [bæt]), taps (e.g. [bæɾ]), and 1-feature place mispronunciation (e.g. [bæp]). All words were presented in the 'now' and 'again' context. Listeners were expected to choose between two written labels on a computer screen – the printed text of the target word (*bat* in all the cases above), and a distractor - a series of "XXX" matched for length to the target word. We analyzed accuracy as well as reaction time data using linear mixed effects models with frequency of the target word as a covariate. Model comparisons without the variable of interest are presented as chi-squares. Unlike in Ranbom *et al*., where listeners were at ceiling (>99%), in Experiment 1 listeners selected the target word more often in the stop (96%) than in the tap condition (82%). This confirms the privileged status of the canonical form in word recognition (Sumner & Samuel, 2005; Sumner & Samuel, 2009; Tucker, 2011). As expected, listeners also selected the target word in stop and tap conditions more frequently than in the mispronunciation condition (27%). Crucially, listeners selected the target word significantly more when a tap-variant was presented in the match context (89%) than in the mismatch context (74%; $\chi^2(1) = 12.89$, $p < 0.001$). Listeners were also faster at recognizing taps in the match (951ms) compared to the mismatch context (1130ms), $\chi^2(1) = 6.35$, $p = 0.01$.

In Experiment 2, listeners (n = 20) were presented with the same stimuli as Experiment 1 but asked to just look at the target word. Eye tracking data (SR Eyelink) was analyzed using growth curve analysis (Mirman *et al.*, 2008). Overall, target fixation proportions did not differ between the stop and tap conditions, $\chi^2(1) = 2.39$, $p = 0.12$, although listeners shifted their gaze to the target faster in the stop condition, $\chi^2(1) = 186.78$, $p < 0.001$, confirming its privileged status. Importantly, within the tap condition, listeners fixated more to the target in the match compared to the mismatch context, $\chi^2(1) = 4.51$, $p = 0.03$.

The results from the lexical decision as well as the eye tracking experiment confirm that context matters in the processing of word-final tap variants. This result poses a problem for Ranbom *et al*., 2009's word recognition model where stop and tap variants are lexically encoded. However, unlike with assimilated variants, word recognition did not fail completely when tap variants were presented out of context. This is problematic for more traditional inference based accounts of spoken word recognition (Gaskell & Marslen-Wilson, 1996). These results are best explained using an optimal-perception account (Mitterer, 2011) where listeners make maximal use of contextual information to compensate for variation. Where taps are the most common variant (70% in match context; Ranbom *et al.*, 2009), listeners readily recover the target word; where taps occur much more rarely (8.5% in mismatch context; Ranbom *et al.*, 2009), listeners are still able to recover the target word but just not as efficiently as in the match context.

References

Gaskell, G., Marslen-Wilson, W. 1996. Phonological variation and inference in lexical access. *JEP: HPP* 22, 144-158 / Gow, D.W. 2003. Feature parsing: Feature cue mapping in spoken word recognition. *Percep. & Psychophysics* 65, 575–590. / Mirman, D., Dixon, J.A., Magnuson, J.S. 2008. Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *J. Mem. Language* 59, 475-494. / Mitterer, H. 2011. The mental lexicon is fully specified: Evidence from eyetracking. *JEP: HPP* 37, 496-513. / Mitterer, H., Blomert, L. 2003. Coping with phonological assimilation in speech perception: Evidence for early compensation. *Percep. & Psychophysics 65*, 956–969. / Otake, T., Yoneyama, K., Cutler, A., van der Lugt, A. 1996. The representation of Japanese moraic nasals. *J. Acoust. Soc. Am*. 100, 3831–3842. / Pitt, M. 2009. How are pronunciation variants of spoken words recognized? A test of generalization to newly learned words. *J. Mem. Language* 61, 19-36. / Ranbom, L., Connine, C., Yudman, E. 2009. Is phonological context always used to recognize variant forms in spoken word recognition? The role of variant frequency and context distribution. *JEP: HPP* 35, 1205-1220. / Sumner, M., Samuel, A. 2005. Perception and representation of regular variation: The case of final /t/. *J. Mem. Language* 52, 322-338. / Sumner, M., Samuel, A. 2009. The role of experience in the processing of cross-dialectal variation. *J. Mem. Language* 60, 487-501. / Tucker, B.V. 2011. The effect of reduction on the processing of flaps and /g/ in isolated words. *J. Phon.* 39, 312-318.

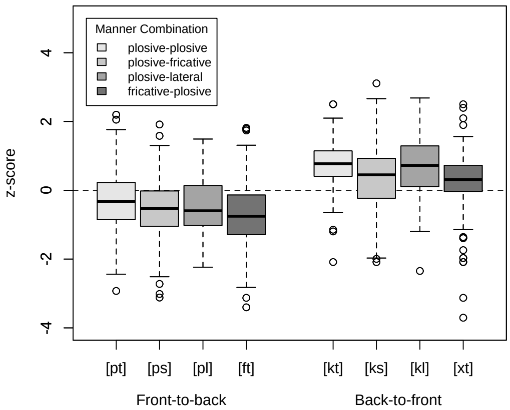
Gestural coordination in word-initial Greek clusters as revealed by kinematic, articulatory, and acoustic data

Jonathan Yip
The University of Hong Kong
yipjonat@hku.hk

Kinematic measures of the relative timing of articulatory movements in the production of consonant clusters in several languages have revealed systematic patterns of articulatory overlap. For example, CC sequences exhibit greater labial-lingual and lingual-lingual gesture overlap (1) for front-to-back clusters (e.g., [bg]) than back-to-front clusters (e.g. [gd]; Chitoran *et al.*, 2002; Chitoran & Goldstein, 2006), and (2) for stop-lateral clusters (e.g., [pl]) than for stop-stop clusters ([pt] or [pn]; Chitoran & Cohn, 2009; Kühnert *et al.*, 2006) and stop-fricative clusters ([ps]; Bombien *et al.*, 2013). The source of these timing patterns is important to theories of speech production, and theories of the relation between production and perception: Is overlap primarily constrained by biomechanical factors, such as constraints on coordination of oral articulators, or by perceptual factors, such as the use of production timing strategies to improve perceptual recoverability and reduce acoustic masking?

This study contributes to current understanding of gestural coordination by examining the influences of order of place of articulation and manner on the timing of constrictions formed by the tongue tip, tongue dorsum, and lips. In this experiment, eight native Greek speakers' productions of word-initial CC sequences [pt ps pl ft kt ks kl xt] were imaged using an ultrasound scanner (lingual data) and video camera (labial data). To assess the patterns of articulatory timing, three types of data were analyzed: articulatory timing, lingual contours, and acoustics. Articulatory timing was measured by tracking aperture along the vocal tract over time for each articulator and identifying time points of gestural achievement and release for that articulator. Degree of temporal overlap between the gestures in each CC sequence was then measured in terms of the duration of lag between the release of C1 constriction and the achievement of C2 constriction (inter-plateau lag). Pooled inter-plateau lag results (Figure 1) showed that dorsal-coronal [kt ks kl xt] were produced with greater intergestural lag than labial-coronal [pt ps pl ft]. However, lag differences across CC manner types varied between speakers, and both biomechanical as well as perceptually-motivated factors are considered.

Figure 1: *z*-scores of inter-plateau lag across sequences (pooled across speakers).



Smooth spline ANOVAs (Davidson, 2006) for lingual contours extracted from gestural peaks in the ultrasound video show between-speaker differences in the degree of anticipatory tongue-tip raising for obstruent-final clusters [pt ps ft kt ks xt] and in the shape of the tongue surface during lateral-final clusters [pl kl]. These differences may provide biomechanically-relevant explanations for the inter-plateau lag patterns observed above. Acoustic details about the strength of plosive release cues and fricative achievement cues as influenced by the degree of potential masking present in each phonetic context are forthcoming.

References

Bombien, L., Mooshammer, C., and Hoole, P. 2013. Articulatory coordination in word-initial clusters of German. *Journal of Phonetics* 41(6), 546–561. / Byrd, D. (1992). Perception of assimilation in consonant clusters: a gestural model. *Phonetica* 49, 1–24. / Chitoran, I., Goldstein, L., and Byrd, D. 2002. Gestural overlap and recoverability: articulatory evidence from Georgian. In Laboratory Phonology 7, C. Gussenhoven & N. Warner (eds.). Berlin: de Gruyter, 419–447. / Chitoran, I., and Goldstein, L. 2006. Testing the phonological status of perceptual recoverability: Articulatory evidence from Georgian. Poster presented at *Laboratory Phonology 10*, Paris, France, June–July 2006. / Davidson, L. 2006. Comparing tongue shapes form ultrasound imaging using smoothing spline analysis of variance. *Journal of the Acoustical Society of America* 120(1), 407–415. / Kühnert, B., Hoole, P., and Mooshammer, C. (2006). Gestural overlap and C-center in selected French consonant clusters. *Proceedings of the 7th International Seminar on Speech Production (ISSP)*, 327–334.

Acoustic differentiation between coda reʈrofle and den al s ops in un a i

Qandeel Hussain, Ivan Yuen

Macquarie University, Australia.

qandeel.hussain@students.mq.edu.au, ivan.yuen@mq.edu.au

Retroflexes typically show F2-F3 convergence near the vowel offset, differentiating retroflexes from dentals (Ohala & Ohala, 2001). However our recent research on Punjabi shows that the formant transitions of retroflexes and dentals in the context of a preceding /iː/ vowel are not different: there is no F2-F3 convergence in this environment. This raises the question as to how the contrast between retroflexes and dentals is maintained in this context. In studies of other languages, retroflexes have been shown to acoustically differ from dentals in having a shorter closure and burst duration (Tabain, 2012; Anderson & Maddieson, 1994). Furthermore, the significant concentration of energy in the spectrum is around 3000 Hz for the retroflexes and 4000 Hz for the dentals (Stevens & Blumstein, 1975). In this paper we show that measures of closure duration, burst duration and Center of Gravity (CoG) of the burst reliably differentiate the retroflex and dental contrast in Punjabi, even in a preceding /iː/ vowel context, where formant transitions do not distinguish the two.

The goal of the present study was therefore to provide a quantitative acoustic analysis of Punjabi voiceless unaspirated retroflex /ʈ/ and dental /t̪/. Ten male Punjabi speakers (22-28 yrs, M=25.2 yrs) participated in an elicited imitation task. They were presented with nine CVːʈ and nine CVːt̪ words, preceded by the long vowels /iː aː uː/, all matched with a picture. The test items were embedded in a pre-recorded Punjabi carrier sentence [keː___əʤ] 'say___today', randomized, and presented in eight different blocks (144 items/speaker). All productions were recorded at 44.1 K with a Zoom digital recorder. The closure and burst durations of the total 1440 test items were then acoustically coded. The cessation of regular pitch pulsing of the preceding vowel was marked as the onset of the closure. Voicing during the closure was not considered part of the vowel. The burst release was labelled as one strong vertical spike in the waveform, signaling the abrupt release of a stop consonant. The CoG of the burst spectrum was measured.

We used GLM repeated-measures ANOVA with the closure duration, burst duration and CoG as dependent variables and the place (retroflex vs. dental) and preceding vowels (/iː/ vs. /aː/ vs. /uː/) as within-subjects factors. As expected, the results revealed that there was a significant effect of place on the closure duration ($F(1,9)=24.029$, $p=.001$), burst duration ($F(1,9)=31.307$, $p<.001$) and CoG ($F(1,9)=42.440$, $p<.001$). In the preceding /iː/ vowel context, the effect of the place on closure duration ($F(1,9)=20.819$, $p=.001$), burst duration ($F(1,9)=20.939$, $p=.001$) and CoG ($F(1,9)=29.387$, $p<.001$) was also significant. This suggests that closure duration, burst duration and CoG can reliably differentiate the retroflexes from dentals in this environment. Similar results were found for the other vowels, with a significant effect of the place on closure duration (/aː/: $F(1,9)=27.677$, $p=.001$; /uː/: $F(1,9)=11.781$, $p=.007$), burst duration (/aː/: $F(1,9)=39.501$, $p<.001$; /uː/: $F(1,9)=30.801$, $p<.001$) and CoG (/aː/: $F(1,9)=20.224$, $p=.001$; /uː/: $F(1,9)=54.639$, $p<.001$). This is shown in Figs. 1-3 below, with the error bars indicating the standard error of the mean.

These findings therefore show that the acoustic cues for the retroflex and dental contrast in Punjabi are robust in the closure duration, burst duration and CoG, and these cues are present in all three vowels. This would therefore help listeners to distinguish retroflex from dental stops, even in the context of a preceding /iː/ vowel, where formant transitions are not distinctive. These results are also consistent with studies of Australian languages showing that apicals (retroflexes, alveolars) have shorter closure and burst duration compared to laminals (dentals, palatals) (Arrernte: Tabain, 2012; Tiwi: Anderson & Maddieson, 1994).

Figure 1: Mean closure duration.

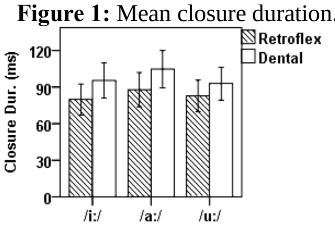


Figure 2: Mean burst duration.

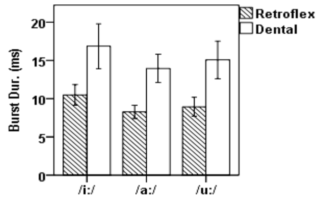
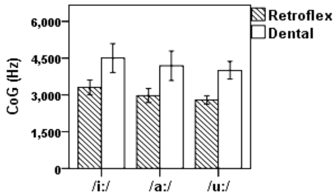


Figure 3: Mean CoG in Hz.



References

Anderson, V. B. & Maddieson, I. 1994. Acoustic characteristics of Tiwi coronal stops. *UCLA Working Papers in Phonetics* 87, 131-162. / Ohala, M. & Ohala, J. J. 2001. Acoustic VC transitions correlate with degree of perceptual confusion of place contrast in Hindi. Travaux du Cercle Linguistique de Copenhague 31, 265-284. / **Tabain, M.** 2012. Jaw movement and coronal stop spectra in Central Arrernte. *Journal of Phonetics* **40,** 551-567. / Stevens, K. N. & Blumstein, S. E. 1975. Quantal aspects of consonant production and perception: a study of retroflex stop consonants. *Journal of Phonetics* 3, 215-233.

Stability in perceiving non-native segmental length contrasts

Yuki Asano and Bettina Braun

Department of Linguistics, University of Konstanz, Germany

(yuki.asano, bettina.braun)@uni-konstanz.de

The discrimination of non-native segments and segmental length contrasts is expected to be difficult for L2 listeners when such contrasts do not exist in the native language (theoretical supports for segmental contrasts see e.g. Best *et al*., 2001). Despite this prediction, non-native listeners have sometimes been reported to discriminate non-native segmental length contrasts better than predicted (e.g. Hayes-Harb and Masuda, 2008 for English L2 learners of Japanese). We postulate that the high sensitivity to these non-native length contrasts may be task-related: if the task only requires auditory processing (instead of phonological processing, cf. Baddeley, 1986; Gerrits, 2001), sensitivity may not be dependent on L1 categories, particularly when listeners are able to concentrate on the contrast in question, i.e., when no other distracting information is interfering.

We tested the limits of the discrimination of Japanese consonantal length contrasts in disyllabic nonsense-words with three groups of listeners (48 German learners of Japanese, 24 German non-learners and 24 Japanese natives) by varying (1) the inter-stimulus-interval (henceforth ISI, 300ms vs. 2500ms) to dissociate acoustic processing from the phonological one and (2) the pitch contours associated with the length contrast (high-flat pitch, vs. high-falling pitch).

We analyzed the participants' sensitivity to consonantal length contrasts by calculating d' scores (Macmillan and Creelman, 2005), normalized by each participants' sensitivity to vocalic length contrasts (which served as a baseline). A linear mixed effects regression model with *d’ scores* as dependent measure and *language group*, *pitch*, and *ISI* as fixed factors and *participant* as a random factor including random slopes for the fixed factors showed a significant three-way interaction ($p < 0.003$). In the flat pitch condition, there was an interaction between *language group* and *ISI* (the non-learners’ d’ scores decreased in the long ISI condition, but not those of the other two groups, $p < 0.03$). In the falling pitch condition, the main effect of *language group* was detected (the natives’ d’ scores were higher, followed by the learners’ ones, $p < 0.01$, then by the non-learners’ ones, $p < 0.01$), see Figure 1.

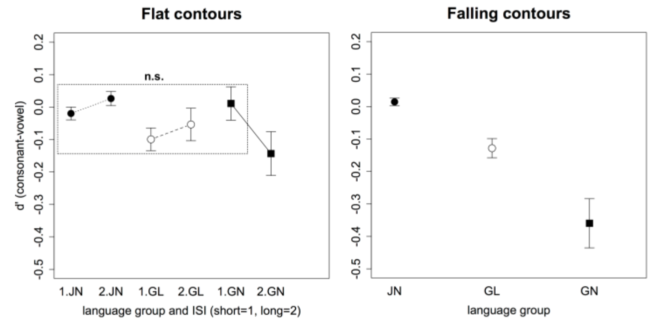


Figure 6: d’ scores for he fla and he falling pi ch

To summarize, the results showed very good discrimination in all groups when task demands were lowest (flat pitch and short ISI). Under these circumstances, even non-learners without any exposure to the L2 could discriminate the non-native consonantal length contrast. The two experimentally induced task difficulties had a differential effect on the two non-native groups. When processing tapped into phonological storage (long ISI), only the non-learners' discrimination abilities decreased. On the other hand, both learners and non-learners showed a decrease in sensitivity in the falling pitch condition compared to the flat pitch condition. The psycho-acoustic complexity of the stimuli hence had a stronger impact on the performance of non-native listeners than the increased memory load (which only affected listeners without prior exposure with Japanese).

Our findings demonstrate that non-native learners have already established categories for representing length contrasts, but that their perception is still vulnerable in less favorable listening conditions with distracting acoustic information. Their reduced sensitivity under increasing task demands appears to be the reason why even advanced learners still face difficulties in remembering length contrasts in more natural learning situations.

References

Baddeley, A. D. 1986. *Working memory*. Oxford: Oxford University Press. / Best, C., McRoberts, G. W., Goodell, E. 2001. Discrimination of nonnative consonant contrasts varying in perceptual assimilation to the listeners native phonological system. *J. Acoust. Soc. Am*. 109, 775-794. / Gerrits, E. 2001. *The categorization of speech sounds by adults and children*. Ph.D. dissertation. Universititeit Utrecht. / Hayes-Harb, R., Masuda, K. 2008. Development of the ability to lexically encode novel l2 phonemic contrasts. *Sec. Lang. Research*. 24, 1, 5-33. / Macmillan, N., Creelman, C. 2005. *Detection theory*. Mahwah, NJ: LEA.

Effect of attention on L2 speech perception: Perception of the Korean three-way stop contrast by English learners

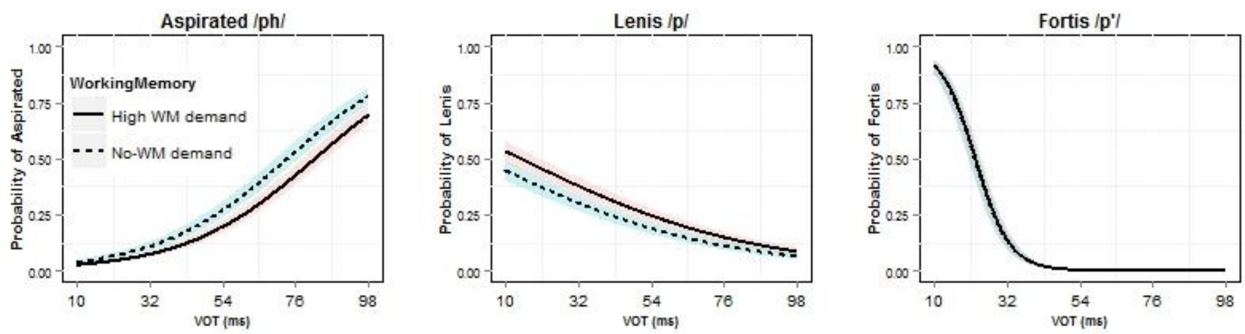
Hyunjung Lee
University of Chicago
hyunjunglee123@gmail.com

Introduction: The phonetic perception of a complex of acoustic cues may differ depend on degree of attention (Gordon, Eberhardt, & Rueckl 1993). Gordon et al. (1993) showed that the effect of VOT to the stop voicing was modulated by attentional demands; under high working memory (WM) demanding condition, English listeners showed a decreased reliance on VOT but an increased reliability on F0 at the onset of the following vowel, the secondary voicing cue. The present study tested the effect of WM on L2 speech perception by examining if and how the perception of the three-way laryngeal contrast among the voiceless stops in Korean (fortis-/p'/, lenis-/p/, aspirated-/p^h/) is modulated by the attentional demands among English speaking L2 learners. In Korean, VOT is the primary cue to distinguish the fortis from the lenis and aspirated stops (fortis < lenis ≤ aspirated), and F0 plays an important role in distinguishing the lenis from the fortis and aspirated stop (lenis < fortis ≤ aspirated).

Methods: Fourteen native English speaking beginning L2 learners of Korean (6 males) participated in this study.. A 30-step continuum (5-step VOT x 6-step F0) was generated based on a /pa/ token produced by a male native Korean speaker. For the dual task paradigm, the phoneme identification task was embedded in a letter recall task. Half of the experimental blocks were conducted for the WM condition; on each trial, the participants were given 5 letters presented one at time on a computer monitor for 1300ms, and asked to type the letters in the order after the identification task. The other half of the blocks was conducted for the No-WM condition without the letter recall distractor. Each participant responded a total of 180 trials (30 stimuli x 2 WM conditions x 3 repetitions).

Results & Discussion: Identification responses were modeled in terms of three binomial mixed-effects logistic regression models (e.g., 1=Fortis, 0=Non-fortis; 1=Lenis, 0=Non-lenis etc.) with three predictors: VOT, F0, and Working Memory (WM, No-WM). Each model maximally included the three predictors and all possible interactions, and the final model was obtained by forward selection. The interaction of WM x VOT was observed for Aspirated (b = -0.61, z-value = -3.13, $p < 0.01$) and Lenis (b = 0.25, z-value = 2.06, $p < 0.05$), but not for Fortis; the interaction of WM x F0 was not observed in any of the analyses. Figure 1 shows the steeper regression curve of Aspirated in the No-WM condition (dotted line) than in WM (solid line); in the No-WM condition where the learners focused only on the identification task, English listeners relied more on VOT compared to the WM condition in predicting Aspirated, which is in line with Gordon et al. (1993) with stop voicing contrast in English. For Lenis, on the other hand, the effect of WM is opposite from Aspirated, showing the steeper curve in WM (solid line) than in No-WM (dotted line). For Fortis, the steepness of the curves does not differ between the two WM conditions. Overall, the effect of attentional demands in the perception of English voicing contrast is consistent for the aspirated stop, but not the other two. The difference in the WM effect might be attributed to the acoustic properties of the three Korean stops. The aspirated stop of which acoustic properties are the most similar to the English voiceless stop (i.e., long VOT, high F0) might have the comparable effect of attention to the perception of L1 voicing contrast, whereas the lenis stop that crucially uses F0 (i.e., the secondary cue to English voicing contrast) might not be affected by the attentional demands in the same way as the similar category (aspirated).

Figure 1: Regression curves as a function of VOT and WM for Aspirated, Lenis and Fortis.



Conclusion: The present findings suggest that under optimal listening environment, the reliance on the primary voicing cue used for the native contrast increases in the L2 speech perception, consistent to L1 perception. But presumably due to L2 learners' phonetic knowledge of voicing, this effect is limited to the phonetic category that is comparable to their native phoneme.

The perception of French reduced speech by native and non-native listeners

Sophie Brand[1] & Mirjam Ernestus[1,2]

[1]Radboud University Nijmegen, The Netherlands

[2]Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

s.brand@let.ru.nl, m.ernestus@let.ru.nl

In casual speech, native speakers do not pronounce all sounds. In French, for example, in words like *repas* 'meal', the first vowel is often absent. Despite the ubiquity of reduction in everyday speech, little is known about how reduced words are perceived by native and non-native listeners.

We investigated whether native and non-native listeners use lexical representations of reduced pronunciation variants for understanding these variants. For *repas*, a French listener's mental lexicon would then contain both /rəpa/ and /rpa/. This hypothesis is in line with research suggesting that French native speakers use representations like /rpa/ in *production* (Bürki, Ernestus and Frauenfelder, 2010) and that native listeners of American English recognize variants through lexical representations in which /nt/ is *replaced* by the nasal flap (Ranbom and Connine, 2007). We focus on the role of lexical word representations with *missing* segments in *perception*.

We conducted a lexical decision task that contained, among many filler words, 44 French words produced either with or without schwa (e.g. 'le repas' or 'le rpas'). Participants had to determine whether the auditorily presented stimulus was a French word or not**.** In French, words differ in how often they occur in a certain pronunciation variant. The schwa in the word *chemise* 'shirt' for instance, is more often absent than the schwa in the word *rebelle* 'rebel' (Racine, 2008). After the lexical decision task, participants indicated the relative frequencies of the two pronunciation variants for each word by choosing a value on a scale of one to six. We examined whether these subjective relative frequencies correlated with speed of recognition of both variants. If so, listeners must have stored these word-specific relative frequencies, which implies that the variants are stored as well.

First, we tested thirty-six French native listeners from Paris. We examined the validity of their frequency estimations by comparing the ratings averaged over listeners to similar ratings obtained by Racine (2008), which correlated well with the frequencies of the variants in a corpus of spoken French. Our ratings correlated well with Racine's participants' ratings ($r = 0.79$, $p < 0.0001$). This finding confirms that the lexicon keeps frequency counts and that there are at least two distinct lexical representations: one for the unreduced pronunciation variant and one for the reduced pronunciation variant. We then investigated whether, in the lexical decision task, the French native listeners reacted faster to reduced and unreduced variants with higher relative frequencies. The results show that this was the case ($p < 0.01$). Apparently, the French native listeners used the lexical representations during perception.

Second, we asked forty-seven Dutch advanced learners of French to perform the same tasks. Their relative frequency ratings showed less variation than those of the French native listeners. Nevertheless, their ratings also correlated with the French participants' ratings obtained by Racine (2008) ($r = 0.46$, $p < 0.01$). This suggests that even advanced Dutch students of French have some knowledge of the relative frequencies of pronunciation variants. Furthermore, their relative frequency ratings correlated with their response times ($p = 0.001$): the non-native listeners reacted faster to variants that occur more often. Importantly, the average rating of the French native listeners did not predict the Dutch participants' response times. Along with the lower correlation with Racine's ratings, this result shows that the frequencies stored by the natives crucially differ from those stored by the non-natives.

These data strongly suggest that both native and non-native lexicons keep frequency counts and that at least two distinct phonological representations in the lexicons are used during perception, one for the reduced variant and one for the unreduced variant. The fact that the French participants' ratings differ from the Dutch participants' ratings and that they therefore do not predict Dutch participants' response times is probably due to the non-natives' limited experience with French casual speech. This may explain why even advanced non-native listeners have problems understanding spontaneous conversations.

References

Bürki, A., Ernestus, M., and Frauenfelder, U. H. 2010. Is there only one "fenêtre" in the production lexicon? On-line evidence on the nature of phonological representations of pronunciation variants for French schwa words. *Journal of Memory and Language*, 62, 421-437. / Racine, I. 2008. *Les effets de l'effacement du Schwa sur la production et la perception de la parole en français*. Thèse de doctorat: Université de Genève, 2008, no. L. 652. / Ranbom, L. J., and Connine, C. M. 2007. Lexical representation of phonological variation in spoken word recognition. *Journal of Memory and Language*, 57, 273–298.

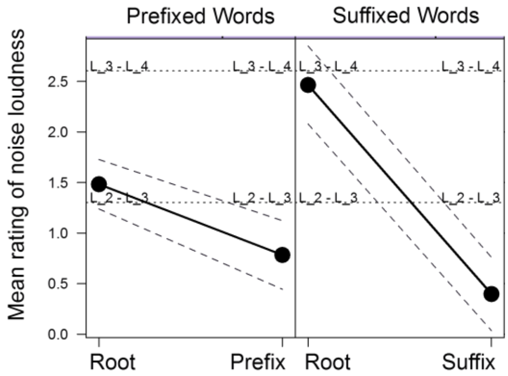
Subjective Perception of Affixation: A Test Case from Spanish

Anne Pycha
University of Wisconsin, Milwaukee
pycha@uwm.edu

Affixation changes our perception of a word, but in ways that we do not fully understand. Typological studies point to the prevalence of suffixation over prefixation as evidence that suffixed words provide perceptual advantages, because they place the root in initial position (Hawkins & Cutler 1988). But this idea is at odds with the experimental literature, which shows that listeners “strip” prefixes from multimorphemic words (Taft & Forster 1975), and provides evidence for a prefixation advantage: the root *pay* and the prefixed word *prepay* facilitate recognition of *payment*, but the suffixed word *payable* inhibits it (Marslen-Wilson et al. 1994; Zwitserlood, & Roelofs 1991, Feldman & Larabee 2001). Sorting out the perceptual consequences of affixation is important for models of word recognition, and could help explain problems in phonology, such as the lack of prefix-triggered alternations on roots (Hyman 2008).

The current study investigates this issue using a noise-rating task in Spanish. Participants heard Spanish phrases that had been partially overlaid with white noise, and assigned a rating indicating how loud the noise sounded, on a scale from 1 (softest) to 5 (loudest). The phrases consisted of verbs plus pronouns; the noise, indicated by strikethrough, coincided with either the prefix (*~~me~~ patea* ‘she kicks me’, *se ~~me~~ pisa* ‘I am stepped on’), the root, or the suffix (*paté~~ame~~,* ‘kick me’, *písa~~me~~lo* ‘step on it for me’). Spanish offers a good test case for our question because the same personal pronoun clitics, which behave phonologically like affixes (Hualde 2005), can precede or follow the verb root, and need not occur at word edges – properties that allow us to focus on affixhood per se. The noise-rating task permits a measurement of listeners’ subjective experience of the word, with the premise that listeners will experience the overlaid noise as softer when the spoken word itself seems perceptually clearer (Jacoby et al. 1988).

A speaker of Colombian Spanish recorded the words. The intensity of each root and affix was calculated separately and white noise was added accordingly at one of three signal-to-noise ratios: +24 dB, +17 dB, or +10 dB. Forty-four native Spanish-speaking participants each heard 36 target phrases, plus 80 fillers, and were asked to attend to phrase meanings during the rating task. Results, analyzed with proportional odds logistic regression, showed an interaction between Noise location and Affix type ($\beta = -1.37$, $t = -7.10$, $p < 0.01$).



Participants rated noise on suffixed roots as very loud, suggesting diminished perceptual clarity, but rated noise on prefixed roots as relatively soft, suggesting perceptual enhancement. These findings confirm the previously reported prefixation advantage, and extend it by suggesting that prefixes facilitate priming of related words precisely because they provide listeners with enhanced perception of the root. Interpreted in light of listener-based theories of diachronic change (Ohala 1993), this finding could help to explain the lack of prefix-root phonological alternations cross-linguistically: just as certain sequences of vowels and consonants encourage misperception more than others, so do certain sequences of morphemes.

References

Feldman, L. B., & Larabee, J. (2001). Morphological facilitation following prefixed but not suffixed primes. *J Exp Psych: 27*(3), 680. / Hawkins, J. A., & Cutler, A. (1988). Psycholinguistic factors in morphological asymmetry. *Explaining language universals*, 280-317. / Hualde, J. I. (2005). *The sounds of Spanish*. Cambridge. / Hyman, L. M. (2008). Directional asymmetries in the morphology and phonology of words. *Linguistics*, *46*(2), 309-350. / Jacoby, L. L., et al. (1988). Memory influences subjective experience: Noise judgments. *J Exp Psych*, *14*(2), 240. / Marslen-Wilson, W., et al. (1994). Morphology & meaning in the English mental lexicon. *Psychological review*, *101*(1), 3 / Ohala, J. J. (1993). The phonetics of sound change. In *Historical Linguistics* (pp. 237-278). London: Longman. / Schriefers, H., Zwitserlood, P., & Roelofs, A. (1991). The identification of morphologically complex spoken words. *J. of Mem & Lang*, *30*(1), 26-47. / Taft, M., & Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. *J. verbal learning & verbal behavior*, *14*(6), 638-647.

The psychological status of the right-branch condition and deaccentuation on *Sais-Sori* in Korean

Hyun Kyung Hwang
National Institute for Japanese Language and Linguistics
hwang@ninjal.ac.jp

Issue *Sais-Sori* phenomenon exhibits tensification of a lax obstruent or lengthening of a nasal consonant in the morpheme initial position of the second morpheme. Tense series are represented by '*'.

(1) E1 E2
mul + koki → mulk*oki 'water + meat → fish'

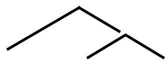
This process has been extensively discussed in the phonological literature, as they are pervasive and complicated in nature. In particular, a number of studies addressed the conditions where Korean *Sais-Sori* applies. One well-known restriction is the right-branch condition. In parallel with rendaku in Japanese, Cook (1991) and Lee (2004) claim that only an element undergoes *Sais-Sori* that is on a right branch in the semantic structure of an ambiguous compound. However, the psychological reality of the conditions remains to be explored through a carefully controlled experimentation. For Japanese rendaku, Kozman (1998) tested the reality of this condition experimentally, and found no conclusive evidence to support the restriction. However, a prosodic factor was not controlled in that study. It is widely recognized that internal structure and accentuation patterns exhibit a strong correlation with *rendaku* (Ito and Mester 2007). Notably similar accentual subordination is observed in *Sais-Sori* as given in (2).

(2) E1 E2 E3
[kjʌ́ul + páta + sóri]
'winter' + 'sea' + 'sound'

a. Right-branch

[[kjʌul p*ada] s*ori]
'sound of winter sea'

b. Left-branch



[kjʌ́ul [páda s*ori]]
'sound of sea in winter'

As shown above, deaccentuation (á→a, ó→o) occurs only in the word reflecting *Sais-Sori* (p→p*, s→s*). Again, this correlation and the role of deaccentuation in the interpretation of compounds were not subject to instrumental investigation. The current study explores this semantic condition and deaccentuation on Korean *Sais-Sori* using novel compounds that have potentially ambiguous structure.

Experiment Six novel compounds made up of 3 real nouns were recorded in 4 conditions; presence or lack of *Sais-Sori* and deaccentuation. *Sais-Sori* and deaccentuation of E2 cue the right-branch structure. E1 ends with a sonorant coda so as to circumvent post-obstruent tensification. E2 was a native Korean word, which can possibly undergo Sais-Sori. In total, 24 stimuli together with 20 fillers were tested. 30 native speakers of standard Korean participated in the two-way forced choice test. Two meanings were provided in full sentences. Participants were asked to choose one of the two possible meanings immediately after hearing each stimulus.

Results Though no categorical dichotomy was found, the presence of *Sais-Sori* and lack of deaccentuation tend to yield the interpretation of right-branch as in (2a). A similar study on rendaku in Japanese (Kozman 1998) reports that the presence of rendaku did not facilitate the right-branch interpretation. However, in that study, the accent patterns for the stimuli with or without rendaku were identical for all items. Thus, it is not clear whether the appropriate accent type that matches to each interpretation was offered. Along the limited role of the two conditions, large variations among the test items were observed.

Implications Beyond the basic observations and impressionistic descriptions of existing patterns, this study investigated the psychological status of the right-branch condition and deaccentuation on *Sais-Sori*. These results can be the basis of comparative studies between *Sai-sios* in Korean and *rendaku* in Japanese. The results also suggest that grammatical factors are not sufficient to explain the many idiosyncrasies observed in the occurrence of the process. The large variation observed in this study highlights the need for an expanded scope of analysis, one which can incorporate the interactivity of the conditions as well as a usage-based approach.

References

Cook, Eun-Do. 1991. Rendaku (Japanese) and sai-sios (Korean): are the similarities fortuitous and spurious?. *Harvard studies in Korean linguistics* IV. 3-12. / Kozman, Tam. 1998. The psychological status of syntactic constraints on rendaku. In Japanese/Korean Linguistics 8, David Silva, (ed.), Stanford: CSLI, 107–120. / Lee, Mun-Sang. 2004. 「韓・日両言語における音韻添加 ： サイッソリ化と連濁・促音化を中心に.」『萩国際大学論集』6(1). 1-19. / Ito, J. and A. Mester. 2007. Prosodic adjunction in Japanese compounds. *MIT Working Papers in Linguistics 55: Formal Approaches to Japanese Linguistics 4*. 97-111. Cambridge: Massachusetts.

Perceptual Relevance of Non-F0 acoustic Correlates in Japanese Accent

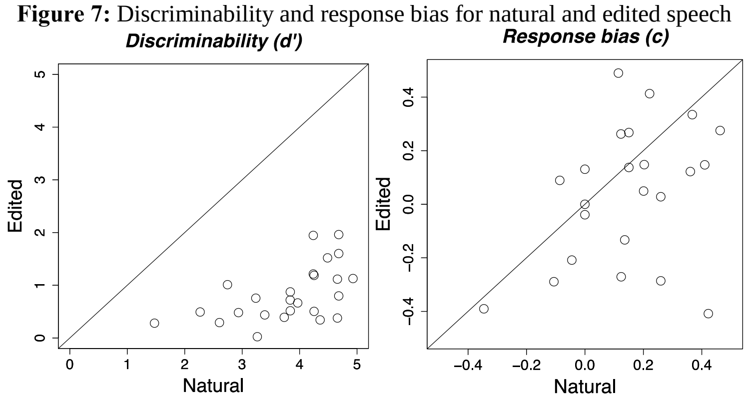
Yukiko Sugiyama
Keio University
sugiyama@hc.st.keio.ac.jp

INTRODUCTION. While prosodic patterns of language are typically realized by multiple acoustic correlates, it is not certain if Japanese accent has acoustic correlates other than fundamental frequency (F0). Past acoustic studies measured duration (Kaiki, Takeda, & Sagisaka, 1992; Beckman, 1986), intensity (Beckman, 1986; Weitzman, 1970), and the rate of devoicing (Yoshida, 2002), but the question remains as to how strong these correlates of accent are. A perception study was conducted using stimuli in which the F0 and its harmonics were artificially removed from natural speech tokens, in order to determine if Japanese accent has correlates other than F0.

METHODS. First, Amano and Kondo (1999) was consulted to thoroughly search minimal pairs of disyllabic final-accented and unaccented words that differ only in the presence or absence of accent. Fourteen minimal pairs were then selected to have both high familiarity ratings and to have unambiguous final or lack of accent. Then, a female native speaker of Tokyo Japanese produced those minimal pairs sentence-medially several times in a random order, of which two tokens of each of the words were used in the perception experiment. Using Praat, F0 and its harmonics were replaced with white noise, while preserving the segmental and timing information of the utterances. Twenty-five native speakers of Tokyo Japanese (18-21 years old) heard eight repetitions of each word in both the original natural speech and its edited version and were asked to choose the word they thought they heard from the two alternatives of the relevant minimal pair. In order to examine the listeners' performance, *d'* scores, a measure of discriminability, and *c* scores, a measure of response bias, were computed.

RESULTS. Figure 1 shows distributions of *d'* scores (left side) and *c* scores (right side) with each circle corresponding to a listener's performance. The x-axes and the y-axes indicate the scores for natural speech and edited speech respectively. As expected, natural speech was identified much more accurately than edited speech, $t(24) = 20.03$, $p < 0.001$. More importantly, when the *d'* scores of edited speech were submitted to a one-sample *t*-test, the result was significantly different from zero, $t(24) = 7.88$, $p < 0.001$; mean = 0.826, indicating that final-accented and unaccented words were identified better than chance. This shows that some acoustic cues other than F0 or its harmonics were present in the signal, which the listeners exploited to distinguish the two types of words. No reliable difference was observed when the *c* scores were analyzed to examine whether the listeners' response bias shifted depending on whether they heard natural speech or edited speech, $t(24) = 1.61$, $p = 0.12$, with natural speech showing a somewhat stronger trend to be identified as unaccented words than edited speech. While past studies report the listeners' tendency to identify unaccented words as final-accented (e.g., Vance, 1995), the listeners in the present study were biased toward unaccented words for both natural and edited speech. In summary, this study is the first to demonstrate in a controlled laboratory setting that non-F0 acoustic correlates exist in Japanese accent and that Japanese listeners can use these correlates to distinguish final-accented and unaccented words.

Figure 7: Discriminability and response bias for natural and edited speech



References

Amano, S., & Kondo, T. 1999. *Nihongo-no goitokusei*. Tokyo: Sanseido. / Beckman, M. E. 1986. *Stress and Non-stress Accent*. Dordrecht, Holland: Foris. / Kaiki, N., Takeda, K., & Sagisaka, Y. 1992. Vowel duration control using linguistic information. *J. IEICE, J75-A*, 467-473. / Vance, T. J. 1995. Final accent *vs.* no accent: utterance-final neutralization in Tokyo Japanese. *J. Phonetics, 23*, 487-499. / Weitzman, R. S. 1970. *Word accent in Japanese*. Detroit, MI: Management information services. / Yoshida, N. 2002. The effects of phonetic environment and vowel devoicing in Japanese. *Kokugogaku*, 34-47, 109.

The productivity and stability of competing generalizations in stress assignment

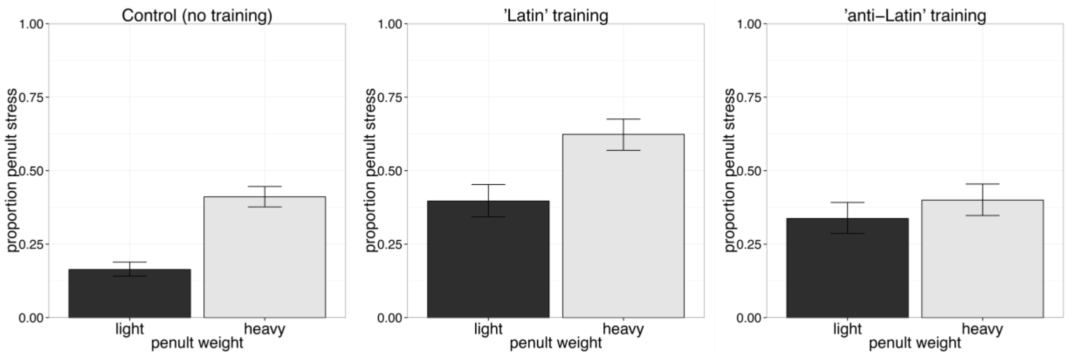
Paul Olejarczuk
University of Oregon
paulo@uoregon.edu

Although English has variable stress that cannot be derived by simple rule, the lexicon contains several productive patterns. One such pattern is trochaic stress (TS), the overwhelming tendency for words to be stressed on the initial syllable (Cutler & Carter, 1987). Another is Latin stress (LS), a weight-sensitive pattern where heavy (CVV or CVC) penults attract stress while light (CV) penults pass it leftward, as in *re.'mem.ber* vs. *'ca.pi.tal* (Hayes, 1995). Studied individually, TS and LS have been shown to affect language processing in a variety of tasks (e.g. Vitevitch et al., 1997; Guion et al., 2003), indicating that speakers infer both patterns from the lexicon. However, very little is known about the relative strength of these inferences or about how they interact when in competition. The present study addressed this gap in two ways. First, a baseline production task examined the relative productivity of TS and LS in nonwords, including cases where both patterns could apply. Second, a learning task assessed the relative stability of the two generalizations by testing the degree to which each could be unlearned.

Seventy-nine American English-speaking adults pronounced trisyllabic nonwords that varied in penult structure between *light* and *heavy*. The words were minimal pairs differing only in the phonotactics of medial clusters, which dictated penult weight (e.g. *ma.da.plazz* vs. *ma.dal.pazz*). The Control group (N=34) performed the task with no training, providing baseline productivity rates of TS and LS. The 'Latin' group (N=22) trained to stress the items according to LS ('stress *heavy* penult, else stress S1'). The 'anti-Latin' group (N=23) trained to stress the same items in the opposite way ('stress *light* penult, else stress S1'). For both training groups, the overall proportion of penult stress was fixed at 50%, in violation of TS. At test, subjects produced new items with light and heavy penults. Stress placement was coded by two coders (Cohen's κ = .933) and analyzed using mixed effects logistic regression, with penult weight and training group as fixed predictors, and maximal random effects structure.

Control subjects exhibited effects of TS (72% overall S1 stress), but penult stress was modulated by weight as expected from LS (41% heavy vs. 16% light, z = 6.3, p < .001). Thus, while both TS and LS generalized to novel words, TS often outcompeted LS (antepenults were usually stressed, even when the penult was heavy). Both training groups unlearned TS by increasing the overall proportion of penult stress, but the Latin subjects came closer to the training distribution (53% vs. 43%, z = -5.1, p < .001). With respect to LS, the Latin group increased heavy penult stress as expected, but surprisingly, they also stressed more *light* penults, roughly maintaining the baseline difference between heavy and light (62% vs. 40%, z = -4.0, p < .001). The anti-Latin group responded to training by increasing the proportion of stressed light penults by about 18%, but remained at baseline for heavy penults. Thus, while TS tends to outcompete LS prior to training, LS is harder to unlearn. The pattern of results may be explained by noting that TS is a simpler, first-order pattern: the placement of stress is not conditional on any characteristics of the word. By contrast, LS is a more complicated, second-order pattern: stress assignment depends on penult weight. Second-order patterns have been shown to be more difficult to learn in the segmental domain (Warker & Dell, 2006), which may make LS harder to acquire from L1 experience, resulting in its lower strength prior to training but greater robustness to unlearning.

Figure 8: Proportion of penult stress by weight and training condition. Bars represent 95% CI.



References

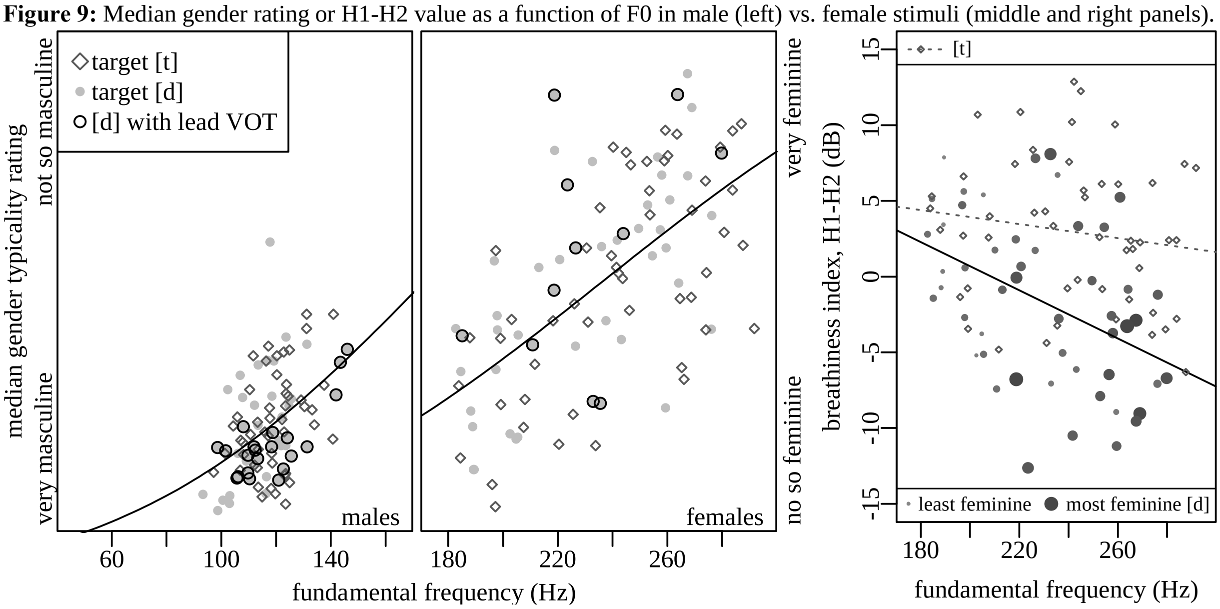
Cutler, A., & Carter, D.M. (1987). The predominance of strong initial syllables in the English vocabulary. *Comp. Sp. & Lang.* 2, 133–42. / Guion, S.G., Clark, J.J., Harada, T., & Wayland, R.P. (2003). Factors affecting stress placement for English nonwords include syllabic structure, lexical class, and stress patterns of phonologically similar words. *Lang. & Sp.* 46, 403–26. / Hayes, B. (1995). *Metrical Stress Theory: Principles and case studies*. Chicago: University of Chicago Press. / Vitevitch, M.S., Luce, P.A., Charles-Luce, J., & Kemmerer, D. (1997). Phonotactics and syllable stress: implications for the processing of spoken nonsense words. *Lang. & Sp.* 40, 47–62. / Warker, J.A., & Dell, G.S. (2006). Speech errors reflect newly learned phonotactic constraints. *JEP: LMC, 32*, 387–98.

Effects of a sound change in progress on gender-marking cues in Japanese

Eun Jong Kong, Kiyoko Yoneyama, Mary E. Beckman
Korea Aerospace University, Daito Bunka University, Ohio State University
ekong@kau.ac.kr, yoneyama@ic.daito.ac.jp, mbeckman@ling.osu.edu

Japanese is often described as contrasting voiced /b, d, g/ to voiceless /p, t, k/. This characterization accords with voice onset time (VOT) values reported in Takada (2011) for speakers from the Kinki region born in the first half of the 20th century, who have consistent voicing lead in /b, d, g/ and short lag in /p, t, k/. However, Takada reports differences in VOT values for /b, d, g/ across subsequent generations of speakers from several dialect regions, including Tokyo. These cross-generation differences, together with results in studies such as Kong et al. (2012), indicate a sound change in progress. That is, younger Tokyo speakers, and especially, younger women, rarely produce a voice bar during the stop closure in words such as *doa* 'door' and they produce words such as *tora* 'tiger' with VOTs that are intermediate between short lag and long lag values. Moreover, regression models improve significantly when other measures besides VOT are included, suggesting the emergence of a "register" system in which lower pitch and a more pressed voice quality at the beginning of the following vowel are beginning to take over the function of lead VOT in differentiating a "devoiced" (or "tense") /b, d, g/ series from a "slightly aspirated" (or "lax") /p, t, k/ series. We evaluated this suggestion by doing a perception experiment in which 20 young native speakers of Tokyo Japanese listened to 200 CV stimuli extracted from words such as *doa* and *tora* produced by 10 men and 10 women. On each trial, the listener either identified the initial stop as "t" or "d" before rating the phoneme category goodness (in one block of trials) or identified the talker as male or female before rating the talker's gender typicality (in the block of trials for the results in Figure 1).

Figure 9: Median gender rating or H1-H2 value as a function of F0 in male (left) vs. female stimuli (middle and right panels).



We use these results to explore how the apparent changes in the phonological function of VOT, pitch, and voice quality interact with the use of laryngeal setting in the socio-indexical marking of gender. For example, women are less likely to produce voicing lead in /b, d, g/. Are they then driving the emergence of a register system? Is the use of other cues more associated with female speech, and is voicing lead taking on a socio-indexical association with male speech? Ohara (2004) and others show that lower pitch and a less breathy voice quality are associated with less feminine styles. The relationships plotted in Figure 1 confirm this correlation between F0 and gender ratings and let us evaluate the local laryngeal effects against the backdrop of the global relationship. Specifically, men's stimuli (left panel) all had low F0 values and were rated as fairly masculine, but stimuli with lead VOT were rated as even more masculine than average. By contrast, women's stimuli showed a large variation in F0 values and correspondingly large ranges of femininity ratings (middle panel) and of H1-H2 values (right panel), with a better separation between lower (more tense) H1-H2 values for [d] (solid line) relative to [t] (dashed line) for stimuli with higher F0 values. That is, more reliable production of the voice quality cue for [d] is associated with stimuli that also have the highest pitch and are rated as most feminine.

References

Kong, E.J, Beckman, M.E., Edwards, J. 2012. Voice onset time is necessary but not always sufficient to describe acquisition of voiced stops: The cases of Greek and Japanese. *J. Phon.* 40, 725-744. / Takada, M. 2011. *Nihongo no gotô heisaon no kenkyuu*. Tokyo: Kuroshio. / Ohara, M. 2004. Prosody and gender in workplace interaction. In: Okamoto, S., Shibamoto Smith, J. S. (eds.), *Japanese language, gender, and ideology: Cultural models and real people*. Oxford: Oxford University Press, 222-239.

Phonological Encoding and Articulation in the Absence of Metrical Spellout

Bryan B. Holbrook

Alan H. Kawamoto

University of California, Santa Cruz
bholbroo@ucsc.edu ahk@ucsc.edu

Stage models of word production, such as the WEAVER model put forth by Levelt, Roelofs, and Meyer (1999), have enjoyed much support. This model assumes word encoding occurs in modular stages such as metrical spellout, segmental spellout, phonological encoding, and finally articulation itself. Results thus far indicate that metrical spellout occurs alongside segmental spellout in parallel, with phonological encoding beginning only after both processes have completed, and articulation following only after phonological encoding has finished (Roelofs & Meyer, 1998). In other words, if the metrical frame of an upcoming word is unknown, phonological encoding of a primed segment cannot begin, and thus no effect of segment priming on acoustic latency would be observed.

However, the results from Roelofs & Meyer (1998) can be questioned for a few reasons. First, participants engaged in an implicit priming task, where word pairs had to be remembered and recalled in addition to produced. The increased task demands of such a paradigm could result in participants choosing accuracy over speed - purposefully delaying acoustic onset to ensure correctness regardless of prime. This would then mask the temporally subtle effects of segment priming, especially in the (presumably) more difficult mixed metrical frame blocks with two, three, or four syllable words tested as opposed to the homogeneous metrical frame blocks which only consisted of two syllable words. In addition, incremental effects are unlikely to be observed when temporal constraints on responding such as a response deadline or explicit instructions to begin early are not presented (F. Ferreira & Swets, 2002; Damian, 2003), neither of which were included in their study.

Since traditional chronometric methodology (i.e., acoustic reaction time of a verbal response) could hide many important effects by focusing on only one point in the entire time-course of a response, we chose to use multiple dependent variables to examine tasks involving speech production. Most relevant, a measure of articulatory latency can be gathered to supplement a traditional acoustic latency measure. Articulatory latency is defined as the onset of labial movement determined through vertical separation of the lips (Kawamoto, Liu, Mura, & Sanchez, 2008). Such a measure is presumably more sensitive to priming effects because of a lower response threshold due to low response "commitment" (easier and less problematic to correct) or less inhibitory control of initial articulatory preparation. Note that this interpretation implies a decoupling of articulatory and acoustic onsets, an effect found by Kawamoto et al. (2008).

The present study utilized such articulatory measures alongside recent advances in methodology to examine the effect of initial segment priming (primed or not primed) on words with varying metrical frames (two or four syllables). Although the number of syllables was unknown, all words were stressed on the second syllable and had a CVC* structure. Measurements of acoustic onset, initial segment duration, and articulatory onset were gathered from recorded video data from a head mounted video camera system. An effect of initial segment priming was found in item analyses (N = 16) for both acoustic latency ($F(1, 12) = 88.306$, $p < 0.001$) and articulatory latency ($F(1, 12) = 53.213$, $p < 0.001$). In addition, some participants occasionally began articulation before the presentation of the complete target, based solely on the prime. Such results imply that phonological encoding, and even articulation itself, can begin without complete knowledge of the metrical frame.

References

Damian, M. 2003. Articulatory duration in single-word speech production. *Journal of Experimental Psychology: Language, Memory, and Cognition, 29*(3), 416-431. / Ferreira, F., & Swets, B. 2002. How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language, 46,* 57-84. / Kawamoto, A. H., Liu, Q., Mura, K., & Sanchez, A. 2008. Articulatory preparation in the delayed naming task. *Journal of Memory and Language, 58,* 347-365. / Levelt, W. J. M., Roelofs, A., & Meyer, A. S. 1999. A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22,* 1-38. / Roelofs, A., & Meyer, A. S. 1998. Metrical structure in planning the production of spoken words. *Journal of Experimental Psychology: Language, Memory, and Cognition, 24*(4), 922-939.

Perception-Production Asymmetry: The Case of Hai-lu Hakka Tone Change and Tone Sandhi

Chia-Hsin Yeh, Yen-Hwei Lin
Michigan State University
chiahsin@msu.edu,liny@msu.edu

In this paper, we provide evidence from tonal processing to shed light on the question of when and in what way speech production and perception exert an influence on phonological grammars in a different manner. We focus on tone sandhi and ongoing tonal changes in Hai-lu Hakka spoken in Taiwan by investigating (i) whether tonal processing errors tend to occur in perception or production or both and (ii) and whether patterns of perception errors are similar to or different from those of production ones.

Kess (1992:63-65) indicates that the classic motor theory and the analysis by synthesis approach support a perception-production interface in which speech perception is mediated by the gestural codes used to produce speech, and so does the psychoacoustic approach by claiming that speech production is guided by the acoustic cues. These interface models suggest that the sensor-motor mechanisms process the same codes/cues but simply in a reverse order. Boersma (1998) and Flemming (2004) propose that speech perception and production are competing forces that shape phonological grammars in an opposite manner: (i) to maximize perceptual distinctiveness and (ii) to minimize articulatory efforts. Based on these previous studies in differentiating perception and production processes, we expect perception-production asymmetry to occur in processing rules and sound changes, but if this is the case, the question is: what are the patterns of this asymmetry?

In the case of ongoing tonal changes (a: low-level→ low-falling, b: high-checked→ high-level, c: low-checked→ low-falling), we recruited 92 Hai-lu Hakka participants in Taiwan, and conducted one production (picture-naming) and three perception (AXB discrimination, tonal identification, and lexical recognition) tasks. The results show that (i) the three affected tones (low-level, high-checked and low-checked tones) are the least accurate categories in all tasks in both perception and production, but (ii) there is an asymmetry in the patterns of production and perception errors: the productions errors reflect the tonal changes (e.g. low-level → low-falling) but the perception errors do not (e.g. low-level→ high-level). In addition, the production errors tend to be faithful to the affected tones in pitch height, whereas the perception errors tend to be faithful in pitch contour (low-level→ high-level, high-checked→ low-checked, low-checked→ high-checked). These findings suggest that Hai-lu Hakka's ongoing changes are production-driven for reducing articulatory efforts since the changing patterns are consistent with the production errors.

In the case of tone sandhi (a. rising-tone sandhi: rising→ low-level in non-final position, b. high-checked-tone sandhi: high-checked → low-checked in non-final position), we recruited 31 Hai-lu Hakka participants in Taiwan, and conducted one production (application of toner sandhi) and three perception tasks. Rising-tone sandhi is phonetically more natural than high-checked-tone sandhi because non-final position (sandhi condition) favors duration-shortening processes (Zhang et al, 2011) and rising to level is duration-shortening but high to low is not. The results show that rising-tone sandhi is more productive and perceptually more confusing than high-checked-tone sandhi: (i) the application rates of rising-tone sandhi are higher and (ii) the perceptual accuracy of rising-tone sandhi is lower. The findings indicate a strong articulatory basis for the the phonetically grounded rule (rising-tone sandhi): ease of articulation leads to greater productivity but tends to cause more perceptual errors as the perceptual distinctiveness is reduced.

Our study demonstrates three common tendencies for the diachronic and synchronic tonal changes in Hai-lu Hakka, and these tendencies may prove to have broader crosslinguistic bases or implications when more studies are conducted on the perception-production asymmetry in tonal processing. First, tonal changes and phonetically grounded tonal rules tend to be production-driven, which is consistent with Xu's (2004) conclusion that tone research needs to take articulatory constraints more seriously. Second, although perception and production forces lead to different sets of error patterns, they both exert an influence on tonal processing. Finally, the three ongoing tonal changes and the more phonetically grounded rule (rising-tone sandhi) have one feature in common: the target tones of the tonal changes and the sandhi tone are all faithful to the affected/underlying tones in pitch height, which may suggest that articulatory dynamic tends to keep pitch height in tonal processing for ease of articulation, but that also tends to cause perceptual confusion.

Selected references

Flemming, E. 2004. Contrast and perceptual distinctiveness. In: Hayes, B., Kirchner, R., Steriade, D. (eds.), *Phonetically-Based Phonology*. Cambridge University Press, 232-276 / Kess, J. 1992. *Psycholinguistics: Psychology, Linguistics, and the Study of Natural Language.* Amsterdam: John Benjamins Publishing Company. / Xu, Y. 2004. Understanding Tone from the Perspective of Production and Perception. *Language and Linguistics* 5(4): 757-797. / Zhang, J., Lai, Y., Sailor, C. 2011. Modeling Taiwanese speakers' knowledge of tone sandhi in reduplication. *Lingua* 121(2): 181-206.

VC coarticulation in Taiwanese and its phonological consequences

Yueh-chin Chang, Feng-fan Hsieh, Yi-cheng Chen
National Tsing Hua University
{ycchang, ffhsieh}@mx.nthu.edu.tw, xyullezx@gmail.com

Introduction: This work is a kinematic study of VC coarticulation in Taiwanese (Southern Min) and the main purpose of this effort is to provide an empirical basis for rime phonotactics, a "unique" phenomenon that is very often, if not always, found in Sinitic languages. More specifically, it has been noted that rimes in Sinitic languages are more subject to phonotactic restrictions than those in non-Sinitic languages (e.g. Korean, Vietnamese and many other "familiar" Indo-European languages). For example, Taiwanese does not have rimes such as */-*up*/, */-*ek*/ or */-*aik*/ whereas there is no such restriction on (tautosyllabic) VC combinations in non-Sinitic languages. In the OT literature, markedness constraints are the de facto standard analysis of these gaps. For example, Chung (1996) invoked OCP-(lab) to rule out rimes such as */-*up*/ in Taiwanese, while RIME HARMONY has been proposed for the generalizations that the segments within a rime must agree in backness and roundedness in Standard Chinese (Duanmu 2007, *a.o.*). Therefore, it is puzzling why effects of RIME HARMONY are hardly evident in non-Sinitic languages (and vice versa).

Hypothesis: Goldstein et al. (2006) and Nam (2009) proposed that anti-phase (sequential triggering) is hypothesized for the VC relation, primarily based on evidence from non-Sinitic languages. Under this view, vowel and coda consonant gestures are mostly *sequentially* rendered, exhibiting lesser degree of coarticulatory overlap. So our working hypothesis is that this crosslinguistic difference may be attributed to the *greater* degree of coarticulation between the nucleus vowel and the coda consonant in Sinitic languages. Phonotactic restrictions emerge once contextual coarticulatory effects, primarily involving assimilation in the tongue body (and the lip position), are phonologized.

Method: We used a Carstens AG500 to record articulatory kinematics in this study. Three (3) male native speakers of Taiwanese were recruited for this experiment. They were aged 20 to 25 years in 2013. All target words have the same CVC syllable structure and similar segmental components. The onset consonant is either /p/, /t/, or /k/ and the coda consonant is /p̚/, /t̚/ or /k̚/; the vowel is either /a/, /i/ or /u/ (and the tones can be either mid or high). The target words were randomized and repeated six (6) times, yielding a total number of 324 tokens (=18 words * 6 times * 3 speakers). Following Iskarous et al. (2010), LE (Locus Equations) slope is based on the X-axis values of Tongue Dorsum (TD) at the midpoint and endpoint of F2 (for VC sequences). Regarding the degree of coarticulation, we also follow Iskarous et al. (2010), using LE slope (obtained via bivariate regression) values to quantify Recasens's (1985) Coarticulatory Resistance (CR): the shallower the LE slope, the greater the CR value.

Results: From Table 1, we see that LE slope values in CV decrease in this order: /p-/ > /k-/ > /t-/ (Sussman et al. 1991). More remarkably, LE slope values in VC are more or less the same as their counterparts in CV. So we were unable to replicate previous results that LE slope values in VC do not differentiate places of articulation (Krull 1988, Modaressi et al. 2004, *i.a.*). Rather, the results of LE slope in VC show that labial stop codas are distinguishable from lingual stop codas. It is also remarkable that the r^2 values are all ≥ 0.84. Finally, we did not include the results measured by means of Recasens's (1985) standard deviation method since these results are by and large consistent with one another.

Table 1 LE slope values in CV vs. VC (based on Tongue Dorsum (TD))

	pV-	tV-	kV-	-Vp	-Vt	-Vk
Spkr 1	0.83, r^2=0.90	0.71, r^2=0.59	0.78, r^2=0.81	1.04, r^2=0.90	0.88, r^2=0.88	0.89, r^2=0.86
Spkr 2	0.96, r^2=0.99	0.52, r^2=0.67	0.80, r^2=0.96	1.03, r^2=0.92	0.82, r^2=0.84	0.74, r^2=0.91
Spkr 3	0.82, r^2=0.97	0.65, r^2=0.59	0.65, r^2=0.76	1.25, r^2=0.93	0.88, r^2=0.91	0.94, r^2=0.88

Discussion: The present results show that i) CV sequences' LE slope values are smaller, suggesting that the onset consonant has greater CR and hence greater aggressiveness towards the vowel; that is, the vowel is coarticulated with the onset consonant to a greater extent than with the coda consonant, ii) nevertheless, there *is* significant degree of coarticulation between the vowel and the stop coda in Taiwanese, if compared with their counterparts in non-Sinitic languages (see also Gao et al. 2011). In sum, we may conclude that in Sinitic languages, the vowel and the coda gestures are *not* necessarily coordinated in the anti-phase mode. Consequently, greater coarticulatory overlap in VC induces (some of) the phonotactic restrictions within the syllable rime. In the same vein, this language-specific gestural pattern may help explain why "falling diphthongs" are not licensed in closed syllables, e.g., */-*aik*/, because postnuclear gestures (/*ik*/ here) *cannot* be coupled sequentially (anti-phase) with the nucleus vowel in Sinitic languages.

Selected References Goldstein, L. et al. 2006. The role of vocal tract gestural action units in understanding the evolution of phonology. In: M. Arbib (ed.) *From Action to Language: The Mirror Neuron System*. CUP, 215-249. / Gao, M. et al. 2011. Intra- and inter-syllabic coordination: An articulatory study of Taiwanese and English. *Proc. 17th ICPhS Hong Kong*, 723-726. / Iskarous, K. et al. 2010. Locus equations are an acoustic expression of articulator synergy. *J. Acoust. Soc. Am.* 128.4, 2021-2032. / Recasens, D. 1985. Coarticulatory patterns and degrees of coarticulatory resistance in Catalan CV sequences. *Language and Speech* 28, 97-114.

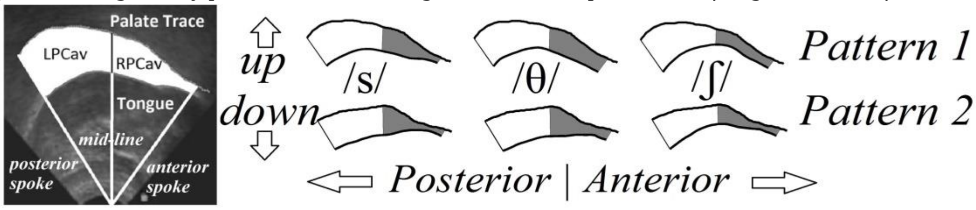
Posterior cavity and aperture distance oppositions in English coronal fricatives

Simon Gonzalez[1], Mark Harvey[1], Michael Proctor[2]
(The University of Newcastle[1], Macquarie University[2])
simon.gonzalez@uon.edu.au, mark.harvey@newcastle.edu.au,
michael.proctor@mq.edu.au

While most studies of English coronal fricatives have focused on the anterior oral cavity, fricative production has been shown to involve complex articulatory-acoustic relationships in all parts of the vocal tract (Catford, 1997; Shadle, 1985). Dart (1991) examined tongue body articulation and found that English speakers can produce /s/ as apical or laminal, and Iskarous et al. (2011) demonstrated that /s/-production involves precise characteristic dorsal trajectories. However, previous studies have not specifically examined whether the geometry of the cavity behind the primary constriction is consistently correlated with different coronal fricatives.

We examined the geometry of the posterior cavity during coronal fricative production by nine female speakers of Australian English. We imaged the vocal tract behind the primary constriction in the mid-sagittal plane using 29.97 fps ultrasound. The ultrasound probe was stabilized using an Articulate Instruments headset. Target segments /θ, s, ʃ/ were elicited in onset position in the words *Thack, sack,* and *shack*. Subjects produced each word five times in the carrier sentence "Please, utter X publically". For each token, the frame corresponding to the point of maximal coronal constriction was identified, and tongue contours were located using EdgeTrak (Li et al. 2003). The data reveal two characteristic articulations (Figure 10): *Pattern 1* - speakers raise the tongue body when producing /s/ and /θ/; *Pattern 2* - speakers lower the tongue body for /s/ and /θ/. All speakers raise the tongue body at the center of articulation of /ʃ/.

Figure 10. *Tongue body patterns characterizing mid-consonant production of English coronal fricatives*



Given the different tongue body patterns, it is possible that there may be no consistent correlations between posterior cavity articulation and segment type for the English coronal fricatives /θ, s, ʃ/. However, we hypothesized that despite the differing tongue configurations, there would be consistent relations between posterior cavity area, posterior cavity aperture distances, and segment type across speakers. To test this hypothesis, following Gick et al. (2006), we measured midsagittal oral cavity area for each speaker between the most anterior point common to the tongue and palate on all images and the most posterior point also common to the tongue and palate on all images ('Posterior Cavity' PCav). We divided this cavity into two portions. The midline division (ML) was located at the lingual apex of the most retracted token of /ʃ/, dividing the total oral cavity into a *Right Posterior Cavity* (RPCav) and a *Left Posterior Cavity* (LPCav). Areas of overall PCav, RPCav, LPCav (mm^2), radial aperture distances (mm) from the tongue contour to the palate trace and area and aperture ratios between segments were calculated: /ʃ/ to /s/, /ʃ/ to /θ/ and /s/ to /θ/. T-tests were used to find significantly different comparisons (Table 1).

Table 1. *Area ratios and Aperture Distance ratios. T-tests with (*) represent comparisons that are significantly different.*

	Area ratios (averages) and t-tests						**Aperture Distance ratios (averages) and t-tests**					
	PCav	***t-test***	***RPCav***	***t-test***	***LPCav***	***t-test***	***pS***	***t-test***	***ML***	***t-test***	***aS***	***t-test***
/s/ vs /θ/	*0.99*	0.191	*1.02*	0.137	*1.03*	0.523	*0.99*	0.524	*1.04*	0.195	*0.68*	0.0001*
/ʃ/ vs /s/	*0.77*	0.0001*	*0.65*	0.0001*	*0.88*	0.007*	*1.10*	0.015*	*0.57*	0.0001*	*0.93*	0.219
/ʃ/ vs /θ/	*0.76*	0.0001*	*0.62*	0.0001*	*0.87*	0.019*	*1.08*	0.939	*0.59*	0.0001*	*0.63*	0.0001*

The results show that the PCav area size distinguishes between /ʃ/ vs /s, θ/, but it does not distinguish between /s/ vs /θ/. Aperture distance distinguishes /ʃ/ vs /s, θ/ on the mid-line, and /θ/ vs /s, ʃ/, at the anterior spoke. The opposition /s/ vs /ʃ/ vs /θ/ is distinguished by either a combination of PCav areas and anterior radial aperture, or a combination of mid-line and anterior radial apertures. Therefore, there are consistent segmental correlations to the dorsal articulation of English coronal fricatives. These data are consistent with the conclusions of Stone and Lundberg (1996) that tongue body has a limited repertoire of configurations which can be recruited during consonant production.

References

Catford, J. C. 1977. *Fundamental Problems in Phonetics,* (Indiana Uni. Press), 154–156 / Dart, S. 1991. Articulatory and Acoustic Properties of Apical and Laminal Articulations. *UCLA Working Papers in Phonetics*, 79. / Gick, B., Campbell, F., Oh, S., Tamburri-Watt. L. 2006. Toward universals in the gestural organization of syllables: A cross-linguistic study of liquids. *Journal of Phonetics*. 34, 49-72. / Iskarous, K., Shadle, C.H., Proctor, M.I. 2011. Articulatory-acoustic kinematics: the production of American English /s/. *J. Acoust Soc. Am*. 129(2), 944-54. / Shadle, C. 1985. *The acoustics of fricative consonants,* MIT-RLE Tech. Report 506. / Stone, M., Lundberg, A. 1996. Three-dimensional tongue surface shapes of English consonants and vowels. *J. Acoust. Soc. Am.* 96(6), 3728-3736.

Are acoustic cues sufficient for the identification and discrimination of Spanish approximants? Evidence from a production study.

Mauricio Figueroa, Bronwen G. Evans

Speech, Hearing & Phonetic Sciences Research Department; University College London
m.figueroa.12@ucl.ac.uk, bronwen.evans@ucl.ac.uk

As in any other variety of Spanish, Chilean Spanish /b/, /d̪/ and /g/ frequently surface as the approximants [β̞] (or [ʋ]), [ð̞] and [ɣ̞] in most phonetic contexts (Cepeda, 1991). Depending on variables such as speech rate or task, these approximant consonants can display degrees of lenition, which ultimately can lead to elision. While it is expected that the phonological reconstruction of the elided units is aided by complementary linguistic and non-linguistic cues (Ernestus, Baayern & Schreuder, 2002; Mitterer & Ernestus, 2006), evidence from a production study for Chilean Spanish indicates that these complementary cues might be necessary to adequately identify and discriminate approximant consonants even when substantial acoustical information is present in the signal.

The production study consisted of recordings from 10 participants (5 males, 5 females) in three tasks: word-lists, short texts and semi-guided conversation. A researcher-led and an automated method of segmentation and extraction of acoustical parameters were applied. Segmentation of approximants is well-known to present a number of difficulties (Turk, Nakai & Sugahara, 2006), and in order to assess which of these two methods provided more reliable acoustic measurements multinomial logistic regression analyses were carried out. Results suggested that the researcher-led method produced more reliable results and these measurements were used in all future analyses.

Overall, the results indicated that duration, intensity and F1 were good correlates of degree of lenition, but not necessarily of phoneme category (especially for the /b/ versus /d̪/ distinction). A series of ANOVA analyses showed significant effects of phoneme category on duration, intensity, F1 and F2, but Bonferroni-corrected *t*-test comparisons showed only significant differences in duration, intensity and F1 between /b/ and /g/, and between /d̪/ and /g/. For F2, there were no significant differences between /d̪/ and /g/. Lastly, within phoneme categories, there were significant differences in duration, intensity and F1 in vocalic, open and closed approximant consonants.

A subsequent perceptual identification experiment investigated how easily listeners were able to identify these tokens. Twenty-nine participants completed an online experiment in which they identified 144 VCV sequences containing approximant variants of /b, d̪, g/, extracted from the production corpus. Overall identification performance was high (average 82.76% correct), but listeners performed differently with different consonants (see Table 1).

Table 1: Confusion matrix for phoneme identification of /b/, /d̪/ and /g/ (chance level at 33%).

	Reference (%)		
Response	/b/	/d̪/	/g/
/b/	81.47	22.07	6.69
/d̪/	4.09	75.86	3.25
/g/	14.44	2.07	90.06

A mixed-model analysis for binomial data with phoneme as a fixed factor and participants as a random factor demonstrated that there was a significant main effect of phoneme, $\chi^2(2) = 33.346$, $p < .001$. These results, which were confirmed by orthogonal contrasts, suggest that /b/ and /d̪/ are significantly more difficult to identify than /g/, and that /d̪/ is more difficult to identify than /b/. An inspection of the confusion matrix showed a relatively strong unidirectional confusion for /b/ perceived as /g/, and a stronger unidirectional confusion for /d̪/ perceived as /b/. Considered together, the results of the production and perception studies suggest that listeners likely use additional non-segmental cues to perceive these consonants, particularly in the case of /d̪/.

In order to determine if listeners do use other cues to achieve identification, two groups of perception experiments are currently in preparation: firstly, experiments to determine how duration, intensity, F1 and F2 are weighted in the perception of /b, d̪, g/ in both full and reduced forms; and secondly, experiments to determine if other non-segmental variables (e.g., stress, word-status) aid the perception of the consonants in both full and reduced forms.

References

Cepeda, G. (1991). *Las consonantes de Valdivia*. Valdivia: Universidad Austral de Chile. / Ernestus, M., Baayen, H., & Schreuder, R. (2002). The recognition of reduced word forms. *Brain and Language*, *81*(1), 162-173. / Mitterer, H., & Ernestus, M. (2006). Listeners recover /t/s that speakers reduce: Evidence from /t/-lenition in Dutch. *Journal of Phonetics*, *34*(1), 73-103. / Turk, A., Nakai, S., & Sugahara, M. (2006). Acoustic segment durations in prosodic research: A practical guide. *Methods in empirical prosody research*, *3*, 1-28.

Acoustic and lingual variation within the 4-way contrast of vowel height in French

Laurianne Georgeton, Tanja Kocjančič Antolík & Cécile Fougeron
Lab. de Phonétique et de Phonologie UMR 7018, CNRS/Univ. Paris 3-Sorbonne Nouvelle, France
[laurianne.georgeton; cécile.fougeron]@univ-paris3.fr; tkocjancic@gmail.com

Phonetic variation in the realization of vowels has been found to depend on the density of phonological inventories (Manuel & Krakow, 1984; Manuel, 1990), with a larger dispersion in the acoustic space of the vocalic targets in a system with fewer vocalic contrasts. In this study, we explore the variability of vowels in the dense system of French, where the four distinctive degrees of height leave little room for phonetic variation. More specifically, we examine how prosodically conditioned variations, earlier attested for peripheral vowels in the language (Loevenbruck 1999, Tabain 2003, Tabain & Perrier 2005), affect height contrasts between front vowels.

The lingual configuration and spectral properties of the four front oral vowels /i, e, ɛ, a/ produced in Intonational phrase-initial position (IPi), where articulatory strengthening is expected (Georgeton & Fougeron, in press), are compared with those in Word-initial position (Wi). Synchronized ultrasound and audio data of four female speakers was recorded, with each vowel being produced in controlled [ip#VC] context and repeated ten times. For acoustic measurements (F1, F2 and F2-F1), three points were taken in the middle of target vowels and then averaged. For lingual configuration, three successive contours were traced manually in Articulate Assistant Advanced in the middle of the vowel and then averaged.

Preliminary results based on the two speakers analyzed so far are illustrated in Figure 1. They show an expansion of the acoustic and articulatory space in IP-initial position, which is more important for one of the speaker (S1). In IP-initial position, the four vowels of the two speakers are more peripheral in terms of F1 for the low vowels /ɛ, a/ and of F2 for /i, e, ɛ/. In terms of lingual configurations, the articulatory space between the maximally contrasting vowels (/i/ and /a/) is strikingly larger in IPi position (solid lines) for speaker S1: there is a slight rising of the front part of the tongue for /i/ (compared to its contour in Wi) and a large lowering of the front part of the tongue accompanied by a slight backing for /a/ (compared to its contour in Wi). Crucially in IPi positions, the difference in tongue contour between vowel pairs of different height is enlarged (see for each pair the difference between the solid lines vs. the difference between the dotted lines) except for the pair /i-e/. The second speaker, S2, shows very little variation in tongue contour according to position. A slight narrowing of the constriction is observed for /e/ and /ɛ/ in IPi, due to a small rising of the tongue, with a slight fronting for /ɛ/. For /i/ and /a/ no clear variations are found.

While two other speakers remain to be analyzed, these results first confirm other observations of speaker-dependent articulatory variations in prosodically strong positions. More interestingly, they show that, when present, articulatory and acoustic variation tends to maximize the contrast between vowels. Furthermore, they suggest that in this dense system of height contrasts, distinctions between the four vowels tend to be maximized by variation in dimensions which are freer to vary (frontedness for /i, e/, lowering for /ɛ, a/).

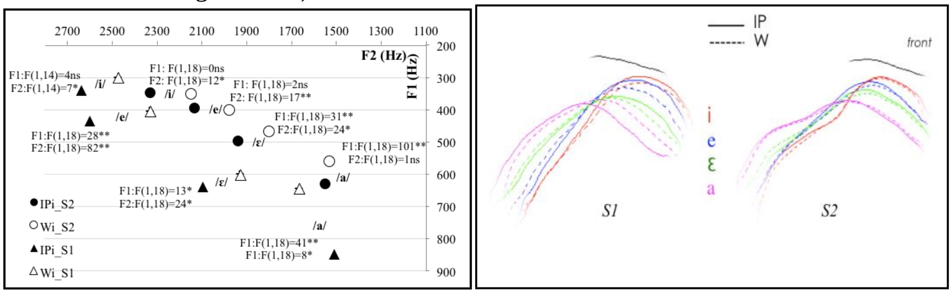


Figure 1 Left: Distribution of the 4 front vowels in the F1/F2 space according to prosodic positions and speakers, with results of the IP vs. W comparisons. Right: Mean tongue contours for the 4 vowels in IPi (plain line) and Wi (dotted line) positions for the 2 speakers.

References

Georgeton, L. & C. Fougeron. in press, Domain-initial strengthening on French vowels and phonological contrasts: evidence from lip articulation and spectral variation. Special Issue, *J. Phonetics* / Loevenbruck H. 1999, An investigation of articulatory correlates of the Accentual Phrase in French, *Proc. 14th ICPhS San Francisco,* 667-670 / Manuel, S.Y. & Krakow, R.A. 1984, Universal and language particular aspects of vowel-to-vowel coarticulation, *Haskins Status Report on Speech Research,* SR-77/78, 69-78 / Manuel, S.Y. 1990, The role of contrast in limiting vowel-to-vowel coarticulation in different languages, *J. Acoust. Soc. Am.*, 88, 1286-98 / Tabain M. 2003, Effects of prosodic boundary on /aC/ sequences : Articulatory results, *J. Acoust. Soc. Am.*, 113 (5), 2834-2849. / Tabain M. & Perrier P. 2005, Articulation and acoustics of /i/ in pre-boundary position in French. *J. Phonetics* 33, 77-100.

Syntactic predictability can facilitate the recognition of casually produced words in connected speech

Malte C. Viebahn[1] *& James M. McQueen*[1,2]
[1]Max Planck Institute for Psycholinguistics, [2]Radboud Universiteit Nijmegen, The Netherlands
malte.viebahn@mpi.nl, j.mcqueen@pwo.ru.nl

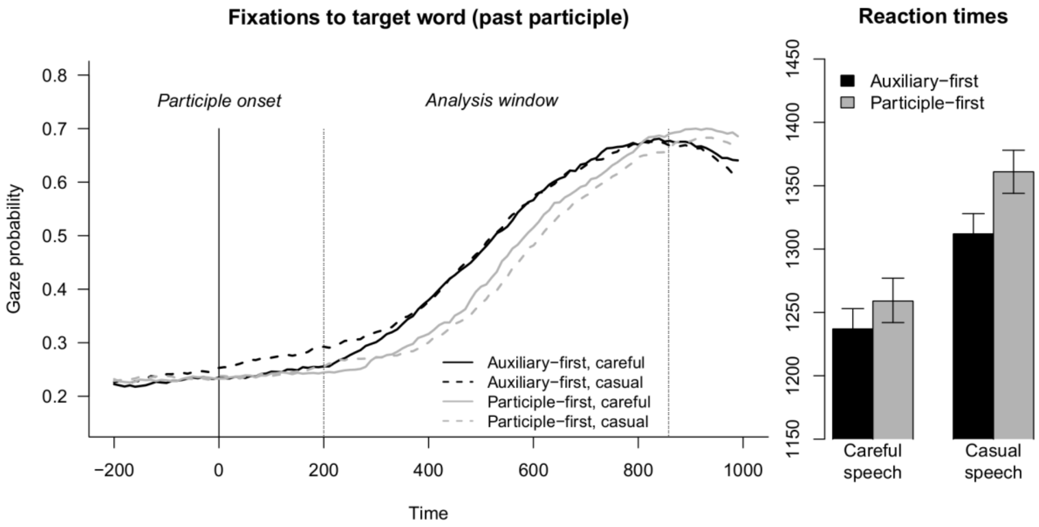
One of the major differences between speech produced in the laboratory and speech that is produced in natural environments is how carefully individual segments are articulated. In spontaneous speech, speakers often omit individual phonemes and even whole syllables (Johnson, 2004), which makes speech recognition more difficult. Listeners may overcome some of these difficulties by making use of information provided by sentential context (e.g. Ernestus, Baayen, & Schreuder, 2002). In the present study, we further investigated the role of contextual information in the recognition of reduced speech. More specifically, we investigated whether the recognition of spoken words is influenced by how predictable they are given the syntactic structure of the sentence in which they occur.

We conducted three eye-tracking experiments that varied in the amount of contextual information provided by the sentence and the time pressure participants were put under. In each experiment, participants heard carefully and casually produced sentences while looking at a visual display containing four printed words: a target word, a phonological competitor, and two distractors. The task of the participants was to select the word which was mentioned in the sentence by clicking on it with a computer mouse. While participants were performing this task, their eye movements were monitored. Syntactic predictability was manipulated by varying the word order of past participles and auxiliary verbs in the stimulus sentences. In Dutch subclauses, past participles can either follow their associated auxiliary verbs (e.g. "Ik weet zeker dat hij heeft geleund op de houten tafel.", *I know for sure that he has leaned against the wooden table*) or they can precede them (e.g. "Ik weet zeker dat hij geleund heeft op de houten tafel.", *I know for sure that he leaned has against the wooden table.*). Participles that follow their auxiliary are more predictable because the auxiliary puts strong constraints on the word classes that can follow it.

The results of all three experiments were similar. Figure 1 shows the fixations to the target word and the reaction times combined across experiments. Participants were more likely to fixate the target word when it was syntactically predictable (β=0.44, t=7.15, p<0.01). Furthermore, listeners recognized careful speech more easily than casual speech (β=0.26, t=4.20, p<0.01). The reaction time data suggest that the disadvantage for casual speech is smaller for the more predictable word order compared to the less predictable one (β=0.023, t=2.50, p<0.01).

Our findings show that syntactic predictability can help in the recognition of spoken words and that this type of information becomes even more useful when processing acoustically reduced speech. This suggests that listeners dynamically adapt to the different sources of linguistic information available to them and by doing so improve their ability to recognize words produced under natural conditions.

Figure 11: Eye fixations (on the left) and reaction times (on the right) from the combined analysis of all three experiments.



References

Ernestus, M., Baayen, H., Schreuder, R. 2002. The recognition of reduced word forms. *Brain and Language, 81*, 162-173. / Johnson, K. 2004. Massive reduction in conversational American English. In: Yoneyama, K., Maekawa, K. (eds.), *Spontaneous Speech: Data and analysis. Proceedings of the 1st Session of the 10th International Symposium*. Tokyo: The National International Institute for Japanese Language, 29-54.

Perceptual attunement to coarticulation: hearing tone in vowel height

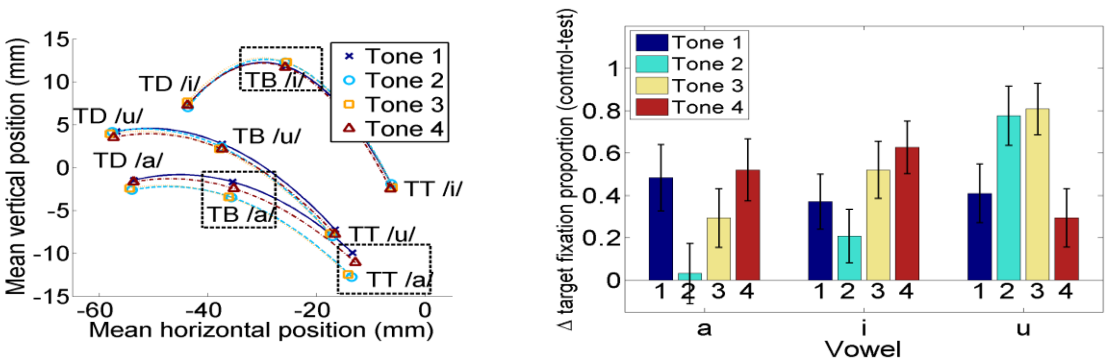
Jason A. Shaw[1], *Wei-rong Chen*[2], *Michael D. Tyler*[1], *Donald Derrick*[1], *Michael Proctor*[3]
[1]University of Western Sydney, [2]National Tsing Hua University, [3]Macquarie University
j.shaw@uws.edu.au, fjueng@yahoo.com.tw, M.Tyler@uws.edu.au, D.Derrick@uws.edu.au, michael.proctor@mq.edu.au

From the perspectives of phonological theory (e.g., Yip, 2002) and of the source-filter theory of speech production (e.g., Fant, 1960), tones and vowels are largely, if not entirely, independent. With rare (and controversial) exception, tones do not condition sound changes or synchronic alternations in vowels. However, a few articulatory studies have found that lexical tone production in Mandarin Chinese exerts an influence on tongue height in vowels, particularly on low vowels (Hoole and Fu, 2004; Erickson, *et al.*, 2004). In complementary production (Electromagnetic Articulography, EMA) and perception (eye-tracking in the visual world paradigm) experiments, we replicate and extend past results on tone-vowel coarticulation and then demonstrate the relevance of these results to speech perception. We find that *the influence of tone on vowel height in production facilitates early recognition of tone in perception*.

Method (production): We recorded six native speakers of Mandarin Chinese (3 male) using the NDI WAVE EMA system. Sensors were attached to the tongue tip (TT), blade (TB), dorsum (TD), lips, jaw, nasion and mastoids. Each speaker produced /pi/, /pa/ and /pu/ syllables with each of the four Mandarin tones: 1 *high*, 2 *low-high*, 3 *low*, and 4 *high-low*. Each syllable was produced 12 times by each speaker, generating a corpus of 864 tokens (12 reps x 3 vowels x 4 tones x 6 speakers). Syllables were presented in Pinyin and randomized with fillers. After computationally correcting for head movement, vowel targets were determined by a 20% threshold of peak velocity of the TD sensor in the opening movement of the vowel. Fig. 1a reports tongue position for each tone-vowel combination.

Method (perception): To investigate the time course of tone perception, we used eye-tracking in a printed-word version of the visual world paradigm. In the test condition, four monosyllabic words differing only in tone, e.g., 八 /pa1/, 拔 /pa2/, 把 /pa3/, 爸 /pa4/, were displayed in Chinese characters while one of the words, e.g., /pa2/, was played over a speaker. Target words combined the four Mandarin tones with three vowels: /i/, /a/, /u/. Sixteen native speakers of Mandarin Chinese were asked to click on the word that they heard, while a Tobii x120 eye-tracker recorded fixations to words displayed in the four corners of the screen. Fixations to the target word in the test condition were compared to fixations in a control condition where competitor words were phonologically dissimilar to the target, e.g., target /pa2/ amongst competitors /ti1/, /t^{h}a3/, /tu4/. Fig. 1b reports difference in fixation proportion to target words across control and test conditions. Smaller differences across conditions (smaller bars) indicate faster tone recognition.

Figure 12: (left panel) Tongue position for tone-vowel combinations. Dotted line rectangles indicate significant effects of tone **(right panel)** Speed of tone recognition across vowels. Smaller bars indicate faster tone identification. Error bars indicate 95% CI.



Results: As indicated by the dotted rectangles in Fig. 1a, significant effects of tone on vowel position were found at the TB sensor for both /i/ and /a/ (and at the TT sensor for /a/ only). Tone 2 lowered the tongue blade for /a/ and raised the tongue blade for /i/. The vowel /u/ was stable across tones. In perception, the speed of tone recognition varied across vowels. For /u/, tones 1 and 4 (tones that start high) were recognized fastest, potentially due to the higher intrinsic F0 of /u/ (Whalen and Levitt, 1995). For /a/ and /i/, tone 2, the tone that exerted a reliable coarticulatory influence on tongue height, was recognized faster than the other tones. Moreover, /a2/, which showed the largest effect in articulation was recognized faster than any other tone-vowel combination.

Discussion: Misattribution of the source of coarticulation is argued to be a major factor in sound change (Ohala, 1981). Although our EMA results show consistent coarticulatory influences of tone on vowel height (for certain vowels), these have not led historically to tone-conditioned vowel alternations. Our perception results indicate consistent attribution of tongue height variation to tone. This ability to perceive tone in tongue height may stem from the intrinsically coextensive nature of tones and vowels and contribute to the apparent resistance of tone-conditioned vowel change.

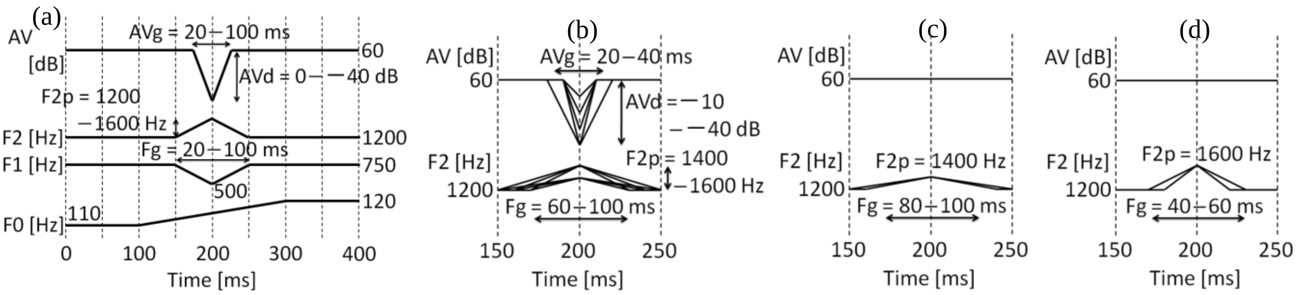
Perceptual Cues of Japanese /r/ Sounds: Formant Transitions vs. Intensity Dip

Takayuki Arai
Sophia University
arai@sophia.ac.jp

Japanese /r/ has several allophones (Kawakami, 1977; Hattori, 1984; Vance, 1987; Saito, 1997), and is typically pronounced as a flap, especially in the intervocalic position. While phrase-initial /r/ is often produced as a plosive-like sound, approximants (including lateral approximants) are another type of allophonic variation that is also frequently observed in real speech. A study of spontaneous Japanese pointed out that the variations were more widely distributed than speech observed in laboratory settings (Arai, 1999). A study of children's speech also reported that the acquisition of /r/ in Japanese is delayed compared with other consonants (Arai, 2013). The same study looked at a wide variety of phones as allophones for real Japanese /r/ sounds: ɖ (retroflex plosive), ɾ (alveolar flap), ɽ (retroflex flap), ɹ (alveolar approximant), l (alveolar lateral approximant), ɭ (retroflex lateral approximant), and ɺ (alveolar lateral flap). The study involved a perceptual experiment, where the likeliness of Japanese /r/ was asked to listeners for each /ara/ stimulus synthesized based on the perceptual cues of formant transitions and a short dip in intensity covering widely distributed sounds in real speech, but only limited discussions were developed.

In the present study, we, therefore, further categorized the experimental results in Arai (2013) related to the allophonic variations in real Japanese /r/. Although the third formant movement plays an important role, especially for the retroflex sounds, Arai (2013) mainly focused on the first (F1) and second (F2) formant transitions and the intensity dip for targeting alveolar sounds. Figure 1(a) shows a schematic plot of the four parameters used to synthesize variations in /ara/ stimuli: depth of the intensity dip, or AVd; duration of the intensity dip, or AVg; total duration of the formant transitions, or Fg; and peak of the F2 frequency, or F2p. The AVd was varied from 0 to -40 dB in 10-dB decreasing steps. The AVg and Fg were independently varied from 20 to 100 ms in 20-ms steps under the condition of Fg >= Ag. The F2p was varied from 1200 to 1600 Hz in 200-Hz steps. A set of 195 /ara/ stimuli in total generated using a formant synthesizer was presented to twenty listeners, who judged the likeliness of Japanese /r/ on a 0-100% scale.

Figure 13: (a) Schematic plot of four parameters as function of time used to synthesize 195 /ara/ stimuli in Arai (2013). (b)-(d) Magnified plots of parameters of stimuli judged as /r/-like sounds with high scores.



In the present study, we categorized the results having high scores of 70% or greater in terms of the likeliness of /r/. The patterns are summarized as diagrams in Figs. 1(b)-(d). In Fig. 1(b), the intensity dip along with the formant transitions yielded a clear alveolar flap sound, which is a typical phone in the intervocalic context. When there is no F2 transition, that is, F2p = 1200 Hz, the resultant stimuli were only judged as Japanese /r/ with low scores. This suggests that the F2 transition towards 1400-1600 Hz acts as crucial cue reflecting the tongue movement against the alveolar ridge. In addition, the intensity dip induces a sensation of the flap sound. In Fig. 1(c), there was no short dip in intensity, but the stimuli sounded an alveolar lateral approximant, and most listeners judged them as a natural Japanese /r/. When the duration of the F2 transition was short enough, such as approximately 50 ms, the resulting stimuli sounded an alveolar lateral flap, which was the case in Fig. 1(d), and also sounded a clear Japanese /r/. In conclusion, the results suggest that continua for real Japanese /r/ have multiple regions where the sounds can be identified as Japanese /r/ and such regions correspond to "sub-phonemes" of real /r/. In other words, Japanese native listeners might perceive such sub-phonemes categorically, and peaks can be observed around the sub-phonemic boundaries in a discrimination test.

References

Arai, T. 2013. *Proc. INTERSPEECH Lyon*. / Arai, T. 1999. A case study of spontaneous speech in Japanese. *Proc. ICPhS San Francisco*, 615-618. / Hattori, S. 1984. *Onsei-gaku*. Iwanami. / Kawakami, S. 1977. *Nihongo Onsei Gaisetsu*, Ohfu. / Saito, Y. 1997. *Nihongo Onsei-gaku Nyumon*. Sanseido. / Vance, T. J. 1987. *An Introduction to Japanese Phonology*. State University of New York Press.

Prosodic strengthening on initial stops in English trochaic *vs.* iambic words

Jiseung Kim, Sahyang Kim, Taehong Cho

University of Michigan, Hongik University, Hanyang University

jiseungk@umich.edu, sahyang@hongik.ac.kr, tcho@hanyang.ac.kr

Previous studies on domain-initial strengthening (henceforth DIS) in English have often been limited to words 1) with initial stress (Fougeron & Keating, 1997; Byrd et al., 2006; Cho & Keating, 2009), which confounds DIS effects with lexical stress, and 2) with voiceless stops, which requires a further understanding of DIS on voiced counterparts. In an effort to fill these gaps in our knowledge of prosodic strengthening, the present acoustic study compares DIS effects on initial segments in iambic (e.g., *<u>pa</u>náche, <u>ba</u>nál*) vs. trochaic (e.g., *<u>pá</u>nel, <u>bá</u>nner*) words with voiceless *vs.* voiced initial segments. It addresses the following questions: (1) to what extent initial segments in unstressed syllables are subject to DIS, and how the effect, if it exits, compares to DIS effects in stressed initial syllables; (2) how phrase-level accent would further interact with DIS effects on lexically stressed vs. unstressed initial segments; and (3) how voiced vs. voiceless initial stops undergo prosodic strengthening in possible interactions among stress, accent, and boundary. If the DIS effect is determined primarily by the segment' proximity to the boundary, quite a robust strengthening effect will be observed even in the absence of lexical stress. Prosodic strengthening effects on voiceless vs. voiced initial stops will be further considered in terms of the phonetic feature enhancement. While voiceless stops in English are produced with longer VOT indicating an enhancement of the phonetic feature {voiceless aspirated}, a critical question is whether strengthening of voiced stops can be seen as an enhancement of {voiced} *vs.* {voiceless unaspirated} or whether it is position-specific.

To this end, target words with iambic stress (*banal, panache, Denise, Tenise*) and trochaic stress (*banner, panel, Daniel, tanner*) were inserted in carrier sentences where position in a phrase (IP-initial vs IP-medial) and phrasal accent (accented vs unaccented) were systematically manipulated. The following shows example carrier sentences with the target word 'panel,' which is underlined. Accented words are upper-cased in bold.

IP-initial, Accented: But after **JOHN** says 'banana', '**<u>PANEL</u>** again' will be the next phrase to say.

IP-initial, Unaccented: But after **JOHN** says 'banana', '<u>panel</u> again' will be the FINAL phrase to say.

IP-medial, Accented: But to say 'banana **<u>PANEL</u>** again' with me is going to be easy.

IP-medial, Unaccented: But to say 'banana <u>panel</u> again' with **ME** is going to be easy.

Eleven native speakers of American English took part in the reading task. Critical acoustic temporal measures for the initial syllables were VOT and Vowel Duration. For IP-medial condition in which closure duration was reliably measurable, two more measures, Closure Duration (henceforth CD) and %-Voicing (during CD) were taken.

Results of RM ANOVAs with Boundary and Accent factors can be summarized as follows. For /p,t/, VOT in stressed initial syllables were longer IP-initially vs. IP-medially only when the target words were unaccented, showing an accent-sensitive DIS effect (Cho & Keating, 2009). No such VOT lengthening effect, however, was observed for /p,t/ in unstressed initial syllables. This suggests that DIS effects are not entirely determined by segments' proximity to the boundary, but they appear to be more robust on the initial segments in stressed syllables (as far as VOT for /p,t/ is concerned). On the other hand, /b,d/ were produced with a short lag VOT IP-initially, but with a negative VOT (voicing lead) IP-medially, regardless of phrasal accent and lexical stress. These results suggest that the direction of the boundary (DIS) effect is towards increasing voicelessness for both voiceless and voiced stops. Although this may be seen as maximizing the syntagmatic (C-V) difference (Fougeron & Keating, 1997), it may also be interpreted as enhancing particular phonetic features IP-initially: **{vl. asp.}** for /p,t/ *vs.* **{vl. unasp.}** for /b,d/. Interestingly, Vowel Duration was shortened (rather than lengthened) IP-initially in all conditions where a significant domain-initial VOT lengthening was found. This indicates some kind of compensatory vowel shortening due to the lengthened VOT.

As for Accent effects, initial vowels were longer under accent even when unstressed. Longer CD was also observed for stops in unstressed initial syllables, showing a leftward spread of accentual lengthening from the following stressed syllable. As predicted, /p,t/ in stressed initial syllables showed longer CD and VOT under accent, enhancing **{vl. asp.}**. Most crucially, regardless of stress, IP-initial /b,d/ with a short lag VOT were produced with even shorter VOT under accent, but IP-medial /b,d/ with a negative VOT was produced with less voicing (shortened voicing lead) rather than more voicing. This was further evident in the increased %-Voicing during CD under accent. This suggests that accent-induced prosodic strengthening on voiced stops /b,d/ is acoustically realized in a way to enhance the language-specific phonetic feature **{vl. unasp.}** rather than **{voiced}** in both IP-initial and IP-medial position. The results will be further discussed in terms of how prosodic strengthening operates on the phonetic feature system of the language.

References

Byrd, D., Krivokapic, J. & Lee, S. 2006. How far, how long: On the temporal scope of prosodic boundary effect. J. Acoust. Soc. Am. 120, 1589-1599. / Cho, T. & Keating, P. 2009. Effects of initial position versus prominence in English. Journal of Phonetics, 37(4), 466-485. / Fougeron, C. & Keating, P. 1997. Articulatory strengthening at edges of prosodic domains. J. Acoust. Soc. Am. 101, 3728-3740.

[DAY 2: Oral Session 5]

Labeling in the Wild: Crowdsourcing versus Categorical Perception

Mark Hasegawa-Johnson, Jennifer Cole, Preethi Jyothi and Lav Varshney

Modern speech technology and speech science depend on large labeled corpora. These large corpora are collected under three assumptions: speech is perceived in terms of discrete categories; there is consistency within and across labelers in the categories that are perceived in a given speech sample; a language can be reliably labeled only by a first-language speaker (and not by non-natives). Machine learning requires labeled corpora, and labeled corpora require technologically literate labelers, therefore any language without a large paid population of such labelers finds itself ignored by speech technology. Recent machine learning theory provides three mechanisms by which we might alleviate the growing speech technology gap between well-resourced and under-resourced languages. First, active learning is a set of methods in which the machine learning algorithm plays an active role in its own education, e.g., by selecting the speech waveforms whose labels would be most informative. In the best case, theoretical guarantees show that active learning can reduce the error rate of a trained classifier from 1/n to exp(-n), where n is the number of labeled data samples. Algorithms for active learning can be tested on well-resourced languages, and having been proven on known tasks in known languages, can then be deployed in order to rapidly develop technology in under-resourced languages. Second, methods of crowdsourcing allow speech labeling tasks to be distributed to a large number of anonymous labelers, some of whom may be less reliable than others. Finally, we can consider the possibility of mismatched crowdsourcing: the acquisition of speech labels from labelers who are not native speakers of the language under study, or by native speakers who lack adequate linguistic training to perform a desired labeling task. Mismatched labeling is a kind of lossy communication channel, according to which some of the information in the original signal has been systematically deleted by the untrained ears of the labeler. As in the case of any other lossy channel, the problem can be solved by a low bit-rate auxiliary channel: in this case, a small amount of data transcribed by a trained expert native speaker may be enough to recover all of the information lost by the mismatched crowdsourcing experiment. In the paradigms of active learning and matched crowdsourcing, theoretical performance guarantees specify upper bounds on the error rates that result from labeling with any given number of labelers. The paradigm of mismatched crowdsourcing has been much less heavily studied, either in theory or in practice, but work in multilingual speech recognition, rapid prosody transcription, and second-language pronunciation error detection suggests the possible shape of future results.

[DAY 2: Oral Session 6]

The private life of stops: VOT in a real-time corpus of spontaneous Glaswegian

Jane Stuart-Smith, Morgan Sonderegger+, Rachel Macdonald*, Thea Knowles+, Tamara Rathcke^*

*University of Glasgow; +McGill University; ^University of Kent

Jane.Stuart-Smith@glasgow.ac.uk; morgan.sonderegger@mcgill.ca; Rachel.Macdonald@glasgow.ac.uk; thea.knowles@gmail.com; T.V.Rathcke@kent.ac.uk

While Voice Onset Time (VOT) is known to be sensitive to a range of phonetic and linguistic factors (e.g. Lisker and Abramson 1967), much less is known about VOT in spontaneous speech, since most studies consider stops in single words, sentences and/or in read speech (but see e.g. Yao 2009). Scottish English is typically said to show less aspirated voiceless stops than other varieties of English, but there is also variation, ranging from unaspirated stops in vernacular speakers to more aspirated stops in Scottish Standard English, and change in the vernacular has been noted (cf Scobbie 2006). Docherty et al (2011) find longer aspiration in younger Scottish speakers, but prefer an explanation of age-grading over sound change. They also show that VOT is subject to subtle sociolinguistic control, indexing local and national identities. But our knowledge of this feature rests on findings from read speech, typically single words from wordlists, even for Docherty et al's (2011) sociolinguistic survey. Nor is this surprising, given that the manual acoustic analysis of large numbers of tokens in naturally-occurring, spontaneous speech is challenging and time consuming.

This paper presents results from a collaborative project which aims to investigate the evidence for possible sound change in progress in stop aspiration and voicing in Glaswegian vernacular, and to develop phonetically-robust automatic methods for obtaining VOT and other characteristics from spontaneous speech recorded outside the phonetics laboratory. The latter is particularly important for assessing this vernacular change, because reading tasks often provoke style-shifting towards the standard. The data are from a real-time Glaswegian corpus covering a hundred years of real- and apparent-time. To tease apart sound change from physiological factors, we analyzed stressed syllable-initial voiced and voiceless stops from a range of recordings from 23 female speakers from three generations (adolescent, middle-aged, elderly), made at two points in time, the 1970s and 2000s, from interviews and casual conversations.

The performance of an automatic measurement algorithm for positive VOT (Sonderegger and Keshet 2012) was refined by manually annotating predictions to identify correct or easily corrected tokens. Overall performance for 11050 stops from 23 speakers was very good, giving 73% measures as correct (58%) or easily corrected (15%); voiceless stops performed even better with 80% judged correct (66%) or correctable (14%). Wrong predictions resulted from issues such as overlapping or noisy speech, or wrong force alignment, but more often (9% overall) from lenited realizations of voiceless and voiced stops. Our semi-automatic method proved very quick (e.g. 40 minutes to annotate 350 tokens).

We first analysed the whole dataset using Linear Mixed Effects modeling with fixed factors of *voicing*, *vowel height* ('high'/'non-high'), *phone* (/b d g p t k/), and *group* (70-Old, 00-Old, 70-Middle, etc) and random intercepts of *speaker*, *word* and *annotator*. We found the voicing contrast to be robust for these speakers, alongside expected lengthening of durations for both denti-alveolar and velar stops in comparison to bilabial stops. Less expected were the interactions of vowel height and place of articulation. For voiceless stops, /p/ showed longer VOT before low vowels, whereas /t/ lengthened before high vowels. For voiced stops, /d/ alone showed lengthening according to vowel height, before high vowels. We also found evidence of real-time change in VOT at all three places of articulation for voiceless stops, with, for example, elderly 2000s speakers showing longer durations than their 1970s counterparts. Apparent-time evidence also occurs in adolescent 1970s speakers using longer VOTs for /b/ and /d/ than elderly 1970s speakers.

Next, and in ongoing work, we added two (significant) prosodic factors, a local speech rate measure (segments/second in each phrase) and position in phrase (initial/non-initial), to model VOT in a subset of 10 speakers. This sharpened the focus of our existing results in three respects: (1) the interaction of vowel height with place became much more significant; (2) for voiceless stops an interaction with *phrase position* with *group* and *phone* indicated different patterns for real- and apparent-time change according to position in phrase, even though phrase-initial stops are relatively infrequent in this corpus; and (3) no evidence emerged to support physiological factors for VOT lengthening.

Our final analysis begins a more detailed investigation of the voiced stops, given that these are only partly described by positive VOT, and we found only a handful of instances of voicing lead. Preliminary results for 5 speakers confirm longer closures for voiceless stops, and substantial proportions of voicing during closure for voiced stops (81%), and even for voiceless ones (48%). We conclude by considering the overall picture obtained across the measures, and the extent to which we are witnessing a real-time phonetic shift in these stop voicing contrasts in our speakers' vernacular.

References

Lisker, L., Abramson, A. S. 1967. Some effects of context on voice onset time in English stops. *Language and Speech*. 10, 1-28. Docherty, G., Watt, D. , Lllamas, C., Hall, D., Nycz, J. 2011. *Variation in Voice Onset Time along the Scottish-English border*. *ICPhS XVII*, Hong Kong. Scobbie, J. M. 2006. Flexibility in the face of incompatible English VOT systems. *Laboratory Phonology*. Berlin, Mouton de Gruyter. 367-392. Sonderegger, M., Keshet, J. 2012. Automatic measurement of voice onset time using discriminative structured prediction. . *J. Acoust. Soc. Am* .132, 3965–3979. Yao, Y. 2009. Understanding VOT variation in spontaneous speech. *Proc. 18th Congress Linguists (CIL XVIII)*. Seoul, South Korea: Korea University.

Articulation rate and VOT in spontaneous speech

Satsuki Nakai[1,2] *& Jim Scobbie*[2]

[1] The University of Glasgow, [2] Queen Margaret University

satsuki@ovod.net, JScobbie@qmu.ac.uk

In many languages Voice Onset Time is a primary acoustic cue that differentiates voiced from voiceless stops in word-initial position (e.g. *bin* vs. *pin*; Lisker and Abramson, 1964). The ranges of VOT values associated with these voiced vs. voiceless stops differ from language to language; therefore, the appropriate range of VOT values for each member of the contrast is learned from their distributions in the ambient language (Maye, et al., 2002).

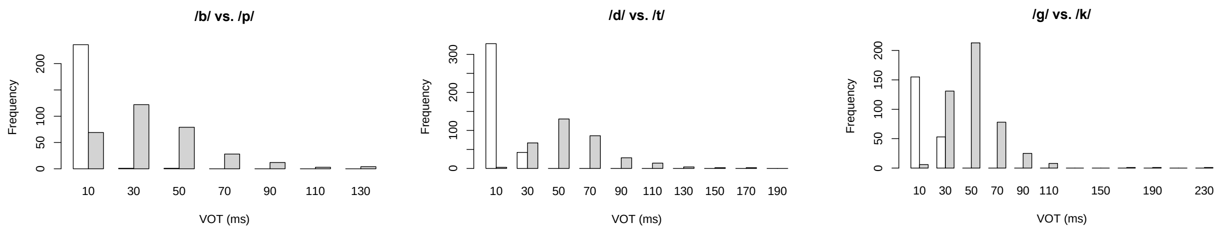
Speech production studies conducted in laboratories suggest that VOT needed to separate voiced from voiceless stops in word-initial position varies considerably depending on the articulation rate (e.g. Miller, et al., 1986). Curiously, in perception studies (e.g. Summerfield, 1981) manipulations of the articulation rate of test materials produce only a very small shift in the VOT category boundary between word-initial voiced vs. voiceless stops. This result is problematic for theories of normalisation in speech perception, which presuppose production-perception links.

Previous production studies on this topic, however, obtained the possible range of VOT by explicitly instructing participants to alter their articulation rate (e.g. to double, halve, and quarter it), and to read test materials. Such methods elicit a range of VOT values that may not be ecologically grounded. In fact, Nagao and de Jong (2007) report a better match between the VOT category boundary in production and perception, when extreme articulation rates are excluded, though they too elicited various articulation rates through explicit instructions.

This study examines the distributions of VOT for word-initial voiced vs. voiceless English stops in spontaneous speech to shed further light on the relationship between articulation-rate normalisation in production and perception. Our hypothesis is that the VOT category boundary between naturally-occurring word-initial voiced vs. voiceless stops is much more stable than laboratory-elicited variation in VOT would suggest, which could explain the small shift in category boundary in perception studies. To test our hypothesis, we examine the VOT of three pairs of word-initial simplex voiced vs. voiceless stops (/b/-/p/, /d/-/t/ and /g/-/k/) in spontaneous speech produced by 12 native English speakers. The speech comes from a BBC radio program entitled “the Lebercht Interview”. Each 44-minute episode features a prominent guest speaker, an artist or administrator in classical music, who talks to the interviewer about their work in a conversational style. VOT is estimated from the audio recordings using Praat.

Preliminary results from 5 guest speakers (1933 stops) reveal that the VOT of each voiced-voiceless pair of word-initial simplex stops in content words are fairly well separated across speakers, and throughout the 44-min conversation (Fig. 1). As a result, a single VOT category boundary for each pair of stops classifies the voicing specifications of these stops rather well: 97% for /b/-/p/, 97% for /d/-/t/ and 95% for /g/-/k/. This level of accuracy is comparable to the accuracy achieved using rate-dependent category boundaries using laboratory-elicited speech data. The classification accuracy further improves when the quality of the following vowel and the presence/absence of lexical stress on the word-initial syllable are taken into account when determining optimal category boundaries. Inclusion of function words (e.g. *to*) undermines classification accuracy.

Figure 1: Distributions of VOT of word-initial stops in content words from 5 speakers (☐: voiced; ■: voiceless).



Thus, as far as word-initial simplex voiced vs. voiceless stops in English content words are concerned, the discrepancy between production and perception literature regarding the magnitude of shift in the VOT category boundary may be attributed to the data collection method used in production studies. VOT values found in spontaneous conversations suggest very little (if any) need for the listeners to shift the VOT category boundary for content words during natural conversations. The findings will be discussed in relation to the cost of normalisation and the predictability of a word.

References

Lisker, L., Abramson, A.S. 1964. A cross-language study of voicing in initial stops: acoustical measurements. *Word* 20, 384–422. / Maye, J., Werker, J.F., Gerken, L.A. 2002. Infant sensitivity to distributional information can affect phonetic discrimination. Cognition 82, 101–11. / Miller, J.L., Green, K.P., Reeves, A. 1986. Speaking rate and segments: A look at the relation between speech production and speech perception for the voicing contrast. *Phonetica* 43, 106–115. / Nagao, K., de Jong, K. 2007. Perceptual rate normalization in naturally produced rate-varied speech. *J. Acoust. Soc. Am*. 121, 2882–2898. / Summerfield, Q. 1981. Articulatory rate and perceptual constancy in phonetic perception. *J. Exp. Psychol. Hum. Percept. Perform.* 7, 1074–1095.

The Syllable as a Prosodic Unit in Japanese Lexical Strata: evidence from text-setting

Rebecca L. Starr, Stephanie S. Shih

National University of Singapore, University of California Berkeley

rstarr@gmail.com, stephsus@gmail.com

While the syllable is considered by many phonologists to be a universal unit in the prosodic structure of language (Selkirk 1982, Kubozono 2003, Hyman 2008, a.o.), the necessity of the syllable in Japanese has provoked considerable debate. Japanese is commonly described as a mora-based language, meaning that the mora, rather than the syllable, is the basic prosodic unit (Vance 1987, Otake et al. 1993, Inaba 1998, Labrune 2012, a.o.). Arguments regarding the necessity of the syllable in Japanese often draw upon evidence from linguistic art forms, including text-setting, the pairing of language and music in song (Poser 1990, Inaba 1998, Labrune 2012). Mora-based segmentation certainly plays a prominent role in Japanese songs, but little notice has been taken of the fact that syllable-based segmentation is also prevalent (cf. Tanaka 2000). We present two studies—corpus and experiment-based—finding that, while mora-based text-setting is the default (i.e., one mora assigned to at least one note), non-moraic variants, in which poly-moraic syllables are undivided and assigned to a single note (e.g., *pa N* → *paN*), are common and accepted in Japanese songs. Our results indicate that this variation is conditioned by both linguistic factors (e.g., lexical stratum) and external factors, including musical style and translation pressures that result from the information density mismatch between English and Japanese (Pellegrino et al. 2011).

Our first analysis examines three corpora: translated Disney songs, translated Christmas songs, and native Japanese *anime* songs. This study focuses on four cases of moras implicated in syllable formation: coda nasal (e.g., *saN*), long vowels (*hoshii*), Vi sequences (*hai*), and high vowels between voiceless consonants (*suki*) (Labrune 2012). Syllabic realizations of each variable were more frequent in the translated corpora ($p < .0001$) and were significantly more likely to occur in foreign stratum words than in native words ($p < 0.0001$), controlling for phonological context. Comparison of the three corpora indicates significant roles for both style and translation factors in predicting non-moraic text-setting.

A follow-up perception experiment was undertaken in order to test whether the patterns found in these corpora are consistent with native listener judgments. Stimuli consisting of synthesized Japanese singing voices were created using the Vocaloid software tool from Yamaha. Listeners rated how well lyrics had been arranged to a given melody for stimuli that used moraic settings, syllabic settings, and misaligned settings that do not match any proposed prosodic unit. Listeners rated both moraic and syllabic settings highly, but rated moraic settings (e.g., *mi-N-to*) significantly higher than syllabic settings (*miN-to*) ($p = .0028$). Listeners also significantly preferred syllabic settings over misaligned settings, including onset-only segmentation (*m-iN-to*) and non-syllabic multi-moraic segmentation (*mi-Nto*) ($p = .0006$). These findings provide positive evidence for the psychological reality of both the mora and the syllable in Japanese prosody.

The evidence presented from our corpus and experimental studies suggest a role for the syllable in all lexical strata of Japanese, regardless of historical origin. We argue that familiarity with English, rather than a difference in underlying prosodic structure, is responsible for a higher frequency of syllabic settings for foreign stratum words. Furthermore, differences in text-setting style across time and genre, along with other idiosyncratic phonological features of sung Japanese, suggest that the dominance of moraic text-setting in certain genres of Japanese song is a prescriptive, stylistic norm rather than a reflection of a lack of syllabic structure.

References

Hyman, L. M. 2008. Universals in Phonology. *The Linguistic Review* 25(1-2), 83 - 137. / Inaba, S. 1998. Moras, Syllables, and Feet in Japanese. *Language, Information and Computation*, 106 - 117. / Kubozono, H. 2003. The syllable as a unit of prosodic organization in Japanese. In: Féry, C.,van de Vijver, R., *The Syllable in Optimality Theory*. Cambridge, U.K.: Cambridge University Press, 99 - 122. / Labrune, L. 2012. Questioning the universality of the syllable: evidence from Japanese. *Phonology* 29:1, 113 - 152. / Otake, T., Hatano, G., Cutler, A., & Mehler, J. 1993. Mora or Syllable? Speech Segmentation in Japanese. *Journal of Memory and Language* 32:2, 258 - 278. / Pellegrino, F., Coupé, C., & Marsico, E. 2011. Across-language perspective on speech information rate. *Language* 87:3, 539 - 558. / Poser, B. 1990. Evidence for foot structure in Japanese. *Language* 66, 78 - 105. / Selkirk, E. 1982. The syllable. In: van del Hulst, H. Smith, N., *The structure of Phonological Representations 2*. Dordrecht: Foris, 337 - 383. / Tanaka, S. 2000. 日本歌謡曲におけるリズムの変遷 [The changing rhythm of Japanese popular songs]. *Proc. of the 14th Meeting of the Japan Phonetics Society*, 153 - 158. / Vance, T. J. 1987. *An Introduction to Japanese Phonology*. Albany, NY: State University of New York Press.

Lexical diffusion of vowel length merger in Seoul Korean: a corpus-based study

Yoonjung Kang; Tae-Jin Yoon; Sungwoo Han

University of Toronto Scarborough; Cheongju University; Inha University

yoonjung.kang@utoronto.ca;tyoon@cju.ac.kr;drysoul@inha.ac.kr

Introduction: In Seoul Korean, long and short vowels contrast in word-initial position (*nun* 'eye' vs. *nu:n* 'snow') but recent studies report loss of length distinction in younger speakers' speech (e.g., Kim 2003, Park 1985, Zhi 1993). The change, however, does not affect all words equally; some words are more resistant to shortening than others. The current study examines the pattern of lexical diffusion, specifically the effect of frequency and phonological contexts on the diffusion of vowel shortening. The results show that while vowels tend to be shorter in high frequency words overall, the sound change proper—i.e., the reduction of contrast over time—is more advanced in low frequency words.

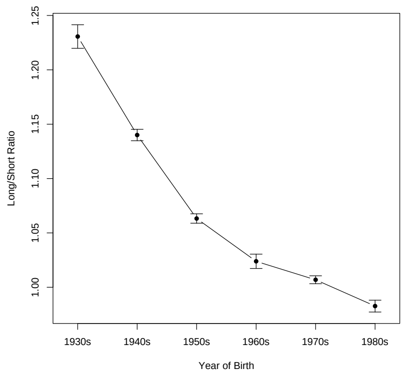
Background: According to Usage-based Phonology (Bybee 2001, Philips 1984, Pierrehumbert 2001), sound changes that involve reduction of articulatory effort tend to affect high-frequency words first while changes that involve (re)analyses affect low frequency words first. While the effect of frequency on lenition in synchronic variation is well documented (Gahl 2008), clear cases of on-going sound change showing the purported frequency effect are sparse and the effect size tends to be minor when compared to that of grammatical conditions (Dinkin 2008, Walker 2012). The present study contributes to this literature by examining the frequency effect in an on-going sound change in Korean.

Data: The data are drawn from *the Reading Style Speech Corpus of Standard Korean* (NIKL 2002). The study examined over 370,000 vowels in word-initial syllable, produced by 118 speakers (59 male and 59 female) stratified for age. The acoustic measurements are aided by the Korean phone aligner developed by the authors. Words are coded for the underlying vowel length (*Great Korean Dictionary*) and the lemma frequency (NIKL frequency list 2007).

Analysis: A mixed-effects model analysis is performed; the dependent variable is the normalized vowel duration (a ratio to the mean short vowel duration of the speaker); the fixed effects include phonological conditions that are known to affect vowel duration (vowel quality, syllable structure, preceding segment, prosodic position, and word length), log lemma frequency (FREQ), underlying vowel length (VL), speaker's gender and year of birth (YOB). To test whether the contrast reduction affects high and low frequency words differently, a three-way interaction of VL*YOB*FREQ is included as fixed effects. Random intercepts of word and speaker are also included.

Results: As **Figure 1** shows, there is a clear trend of contrast reduction between long and short vowels; speakers born in the 1930s exhibit a mean ratio of 1.23 between short and long vowels while the ratio falls below 1 for speakers born in the 1980s. The mixed-effects model shows that all phonological factors show the significant effects in the expected direction (not shown). The main effect of FREQ is significant (β=-0.028,s.e.=0.003, $p<0.0001$); vowels are shorter in high frequency words. The interaction of YOB * VL is significant (β=-0.004,s.e.=0.0003, $p<0.0001$): vowel contrast reduces with younger speakers. The three-way interaction of YOB * VL * FREQ is also significant (β=-0.0005,s.e.=0.00005, $p<0.0001$): vowel contrast reduction in younger speakers is expedited for low frequency words and mitigated for high frequency words. In other words, while high frequency words are in general shorter than low frequency words, the contrast reduction over time, i.e., the sound change itself, affects low frequency words first.

Figure 1:Long/Short Ratio by decade of birth



Discussion: The emerging picture of the frequency effect is a complex one. While as a synchronic pattern, high frequency words are produced with shorter duration than low frequency words in line with frequency effect predicted for a lenition process (cf. Gahl 2008), the opposite frequency effect emerges in diachronic change; frequent words are more resistant to shortening sound change. This diachronic frequency effect suggests that the mechanism underlying the vowel length contrast neutralization may be analytical not phonetic; the high frequency words have more robust representation and are more stable against reanalysis by analogy to the majority pattern (i.e., short vowel). The current finding is significant in that it demonstrates that synchronic frequency effect in lenition may not necessarily lead to the same frequency effect in diachronic sound changes.

References

Bybee, J. 2001. *Phonology and language use.* CUP./ Dinkin, A. 2008. The real effect of word frequency on phonetic variation. *Penn Working Papers in Linguistics 14,1, 97-106.* Gahl, S. 2008. "Thyme" and "time" are not homophones. *Language* 84(3), 474-496. /Kim, S. 2003. *Survey of standard language usage II* [In Korean]. The National Institute of the Korean Language [NIKL]./Phillips, B. 1984. Word frequency and the actuation of sound change. *Language 60, 320-342.*/ Walker, J. 2012. Form, function and frequency in phonological variation. *Language Variation and Change* 24(3): 397-414.

[DAY 2: Oral Session 7]

The peril of sounding manly: A look at vocal characteristics of lawyers before the United States Supreme Court

Alan C. L. Yu[*], *Daniel Chen*[♣], *Katie Franich*[*], *Jacob Phillips*[*], *Betsy Pillion*[*], *Yiding Hao*[*], *Zhigang Yin*[*♦]
University of Chicago[*], ETH Zurich[♣], Chinese Academy of Social Sciences[♦]
aclyu@uchicago.edu, daniel.li.chen@gmail.com, kfranich@gmail.com, jacob.phillips09@gmail.com, betsyschmetsy@gmail.com, yiding@uchicago.edu, yin.zhigang@163.com

Introduction: Individuals make use of many aspects of the speech signals to construct personas and to project hidden desires to the external world. Of interest here is whether vocal characteristics and the perceptual evaluation of them exert an influence on listener behavior. With the exception of a few pioneering studies (e.g., Purnell et al. 1999), this question has remained largely unexplored. In the present study, we examine the vocal characteristics of lawyers arguing in front of the Supreme Court of the United States and link this data to the lawyers' actual win rates in the Court. We show that perceived attributes of voices predict Supreme Court wins, suggesting potential differential labor market treatment of lawyers with certain mutable characteristics such as sounding more or less masculine or confident.

Part I: Methods: In order to obtain listener evaluation of talker voices, we extracted sixty sound clips of male lawyers' Supreme Court oral argument's introductory sentence, which is identical across cases (i.e. "Mister Chief Justice, may it please the court") from the Oyez Project website (http://www.oyez.org/), which is a multimedia archive at the Chicago-Kent College of Law devoted to the Supreme Court of the United States and its work since the installation of a recording system in October 1955. The recordings and the associated transcripts were made available to the public in electronically downloadable format. 200 participants recruited on Amazon's Mechanical Turk listened to the sixty sound clips (normalized for intensity) and were asked to rate the voice sample in terms of masculinity, attractiveness, confidence, intelligence, trustworthiness, and educatedness, as well as the probability of win, on a 7-point scale. **Results:** To understand the relative importance of these seven scalar factors in predicting actual court outcomes, and to avoid problems of collinearity, we conducted binary partitioning of the data using a Classification and Regression Tree (CART) analysis (Breiman et al. 1984) with the *rpart* function in *R* with court outcome (win vs. lose) as the dependent variable and the listeners' responses on the seven scales as predictors. Only two factors, Masculinity and Confidence, remained in the final pruned tree. In particular, individuals with higher masculinity rating are more likely to **lose**, while individuals with high confidence rating are more likely to **win**.

Part II: Methods: To explore the acoustic features that index perceived masculinity and confidence, we measure the following phonetic attributes: formant frequencies (F1, F2) for five stressed vowels (/i, ɪ, ej, ɔ, ʌ/, formant dispersion (average vowel distance from a central point per talker per sound clip), spectral tilt (H1-H2, H1-A1, H1-A2, H1-A3), center for gravity and peak frequency of /s/, speaking rate (phonemes per second), and rhythm (Pairwise Variability Index; Low et al. 2000). **Results:** We conducted binary partitioning of the data using a Classification and Regression Tree (CART) analysis with by-subject normalized perceived masculinity and perceived confidence ratings as separate dependent variables and the phonetic factors as the predictor variables. For perceived masculinity, only six acoustic predictors (F1 and F2 of /ɪ/ and /ʌ/, F2 of /i/, and the center of gravity of /s/) remain in the final pruned tree. Individuals with lower F2 (i.e. more back) for /i/ and /ɪ/ are rated as more masculine, while individuals with higher F1 for /ʌ/ (i.e. lower in vowel height) are rated as less masculine. Finally, very low spectral center of gravity for /s/ is also rated less masculine. In terms of perceived confidence, lower F2 of /i/ and higher PVI (more rhythmic speaking style) are perceived as more confident, but lower vowel height for non-low vowels (i, ɪ, ey) and breathier voices (H1-H2 and H1-A3) are all signs of low perceived confidence.

Conclusion: Our data linking actual outcomes with perceptions, holding constant the words and omitting visual cues, form an uncommon dataset. We show that perceived masculinity and confidence constitute significant predictors of Supreme Court decisions. In particular, more masculine sounding voices are less likely to win a case, while more confidence-sounding voices show the opposite correlation. While the mechanism underlying these connections is still under investigation, it suggests that the perceived vocal characteristics might have significant impact on real world outcomes.

References

Breiman, L., Friedman, J. H., Olsen, R. A., and Stone, C. J. (1984). *Classification and Regression Trees*. Belmont, Wadsworth / Low, E. L., Grabe, E., and Nolan, F. (2000). Quantitative characterization of speech rhythm: syllable-timing in Singapore English. *Language and Speech* 43: 377-401. / Purnell, T., Idsardi, W., and Baugh, J. (1999). Perceptual and phonetic experiments on American English dialect identification. *Journal of Language and Social Psychology* 18(10): 10-30.

Stereotypes predict memory effects for voices

Grant McGuire and Molly Babel

University of California, Santa Cruz and University of British Columbia

gmcguir1@ucsc.edu and molly.babel@ubc.ca

Do listeners selectively process different voices? We present a first step in understanding whether preferential processing of voices affects phonetic encoding by examining whether listeners exhibit improved performance in remembering particular voice types. Some theories predict that listeners should have better voice memory for less typical sounding voices. For example, Goldinger (1998) suggested that atypical tokens are stored in less dense memory space, with thus fewer competitors to inhibit recall. Another possibility is that because atypical tokens are more effortful to process due to energetic masking, greater attentional resources are allocated to signal processing, resulting in more detailed encoding (Clopper, 2012; Mattys et al., 2009). Coming from another direction, research using form-based priming would predict improved performance on standard, more typical-sounding voices (Sumner & Samuel, 2009). Following the dual-processing framework laid out in Sumner et al. (2014), tokens that are preferentially encoded based on higher social value might be more accessible in memory.

We examine these competing predictions by using a voice recall paradigm using 60 voices (30 female) that were previously assessed for explicit, social preference measures of stereotypicality and attractiveness (both measured on Likert scales), as well as the implicit, online processing measures of categorization fluency and intelligibility (both measured as response time). We presented listeners (n = 39, across 4 counter-balanced groups) with an exposure block where they were presented with single words produced by 30 talkers in a blocked-by-talker format. Listeners' task was to type in the words as they were presented. They were then given a surprise memory task where they were presented with the full set of 60 voices and asked to identify the voices as New (not heard before) or Old (heard in the previous exposure block). All voices and listeners were native speakers of American English.

Analyses were approached from the perspective of the listeners and the voices. To model listeners' judgments mixed effects logistic regression models were used to analyze the probability that listeners could correctly identify the voices as new or old. In each model New/Old was entered as a fixed effect, with random slopes and intercepts for Listener. Due to collinearity among the four factors, separate models were calculated using stereotypicality, attractiveness, categorization fluency, and intelligibility as predictors. Only stereotypicality and attractiveness were significant. The stereotypicality model found main effects for Stereotypicality [B = -0.74, SE = 0.09, $p < 0.001$] and New/Old [-5.98, SE = 0.77, $p < 0.001$], and a significant interaction between the two [B = 0.95, SE = 0.12, $p < 0.001$]. The model with attractiveness returned similar results, though the model with stereotypicality as a predictor was a better fit ($p < 0.001$). The interaction between stereotypicality and New/Old showed that new voices which were less stereotypical had higher accuracy rates, and old voices that were more stereotypical had higher accuracy rates.

To model memory with a focus on talker voices, the signal detection theory measures of *d'* (sensitivity) and *c* (bias) were calculated across listeners for each voice. These were used as the dependent measures in linear regression models where each parameter was entered as an independent variable along with talker gender. Voices which were lower in attractiveness [B = -0.29, SE = 0.08, $p < 0.001$; model fit: $F(3, 56) = 4.86$, $p < 0.01$, $R^2 = 0.16$] and stereotypicality [B = -0.36, SE = 0.09, $p < 0.001$; model fit: $F(3,56) = 7.27$, $p < 0.001$, $R^2 = 0.24$] had higher *d'* values—i.e. listeners showed improved performance with such voices. For the intelligibility model, there was an effect of talker gender [B = -5.71, SE = 2.23, $p < 0.05$] and an interaction between intelligibility and talker gender [B = 0.009, SE = 0.004, $p < 0.05$; model fit: $F(3, 56) = 2.92$, $p < 0.05$, $R^2 = 0.09$]; there was improved performance with the less intelligible male voices. The bias results indicated a bias towards reporting attractive [B = -0.2, SE = 0.06, $p < 0.01$; model fit: $F(3, 56) = 6.4$, $p < 0.001$, $R^2 = 0.21$] and stereotypical [B = -0.29, SE = 0.07, $p < 0.001$; model fit: $F(3,56) = 11.6$, $p < 0.001$, $R^2 = 0.35$] voices as old.

Listeners are more sensitive to less stereotypical voices and there is a bias to respond old to more stereotypical voices. These results provide some support for theories where less typical voices are either encoded in less dense space or are, by necessity, given more attention during speech processing. However, the successful voice measure was a based on the subjective concept of stereotype; objective measures of typicality failed to significantly predict the memory patterns. Our finding that stereotypicality best predicts recall suggests that accounts like Sumner et al. (2014)'s are necessary to contextualize the subjective encoding of voice types. With this knowledge the next step is to determine whether speech processing which is more linguistic in nature is similarly affected by voice type.

References

Clopper, C. 2012. Effects of dialect variation on the semantic predictability benefit. *Lang. and Cog Proc*,*27*,1002-1020./Goldinger, S. 1998. Echoes of echoes? An episodic theory of lexical access. *Psych. Review*, *105*(2), 251./Mattys, S, Brooks, J., & Cooke, M. 2009. Recognizing speech under a processing load: Dissociating energetic from informational factors. *Cog Psych*, *59*, 203-243./Sumner, M., Kim, S., King, E., & McGowan, K. 2014. The socially weighted encoding of spoken words: a dual-route approach to speech perception. *Frontiers in Psych*, *4*, 1015./Sumner, M., & Samuel, A. 2009. The effect of experience on the perception and representation of dialect variants. *JML*, *60*, 487-501.

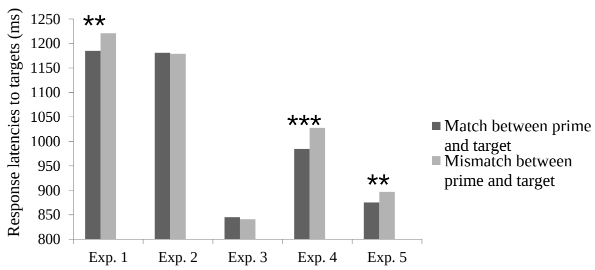
The role of exemplars in speech comprehension

Annika Nijveld, Martijn Bentum & Louis ten Bosch
Radboud University Nijmegen, the Netherlands
a.nijveld@let.ru.nl, bentummartijn@gmail.com, l.tenbosch@let.ru.nl

From an emerging body of research on spontaneous speech it has become clear that spontaneous speech, compared to speech recorded in the laboratory, contains a large amount of variation in how are words are pronounced. The question is how listeners cope with this variation. Abstractionist approaches view variability in the speech signal as "noise" that is irrelevant for further processing. Nevertheless, various studies reported evidence that variability may play a role in speech comprehension. In these experiments, words were repeated and the recognition of the second occurrence of a word was found to be faster when surface details (e.g., speaker voice) of the first (prime) and second occurrence (target) matched compared to when they did not match. Apparently, the surface details of the prime were retained in memory, presumably in the form of exemplars (detailed token representations). Importantly, other studies were unable to replicate this result and under which circumstances exemplars play a role in speech comprehension is still unclear. The presence of exemplar effects is claimed to be related to processing latencies (time-course hypothesis, McLennan and Luce, 2005): exemplar effects arising from indexical variation (i.e., variation in speaker voice and speech rate) can be found when processing is slow. Another possibility is that the effects occur when participants are encouraged to use their episodic memory (e.g., when a large proportion of words is repeated in the experiment) (Hanique, Aalders and Ernestus, 2014). The precise role of exemplars has important consequences for models of speech comprehension; we therefore further investigated this issue.

We conducted two priming experiments, one in which participants performed an old/new judgment task (which requires the use of episodic memory) (Experiment 1) and one in which they performed a semantic categorization task (which does not) (Experiment 2). A prime and target could be spoken by the same or by a different speaker. We also conducted three lexical decision experime[...] nish (Experiment 4) and Dutch (Experiment 5) learners of Engl[...] prime and target could differ in speech style (casual or careful).

Figure 1: Res[...] xperiments 1-5



The results are presented in Figure 1. Within the variation types (speaker voice and pronunciation style, i.e., within Experiments 1-2 and within Experiments 3-5), the time-course hypothesis might hold: the experiments with longer latencies within the two series show exemplar effects. This hypothesis can however not explain our results combined over the series: Experiments 5 and 6 show exemplar effects while they elicited shorter latencies than Experiment 2, which does not show the effect. Moreover, the results of Experiments 1 and 2 show that these two experiments elicited approximately the same mean response latencies, but that only Experiment 1 shows exemplar effects.

The critical factor that is shared across the experiments that show exemplar effects is that in these experiments, participants performed tasks that encouraged them to use their episodic memories. This is in line with the proposal by Hanique, Aalders and Ernestus (2014). That is, the task in Experiment 1 (old/new judgment), but not in 2 (semantic categorization), instructed participants to think back to the previous occurrence of the word. For Experiments 3-5, the non-natives (Experiments 4-5) but not the natives (Experiment 3) show exemplar effects. This is because a lexical decision task poses very different demands on participants depending on their language background. Non-natives might be less confident than natives: the fact that a word does not occur in their L2 lexicon does not mean it does not exist in the language, and, if a word is represented in their lexicon, its representation might be weak. We believe that non-native listeners consequently rely more on their episodic memory than native listeners in a task such as lexical decision. For this reason, they exhibit exemplar effects. Our account implies that exemplars are not part of the mental lexicon, and suggests that the role that exemplars play in the comprehension of spontaneous speech is fairly limited.

References

Hanique, I, Aalders, E., Ernestus, M. 2014. How robust are exemplar effects? *The mental lexicon* 8:3, 269-294. / McLennan, C., Luce, P. 2005. Examining the time course of indexical specificity effects in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition,* 31(2), 306-321.

Effects of phonological neighborhood density on phonetic variation: The curious case of French

Yao Yao[1], *Christine Meunier*[2]

[1]Hong Kong Polytechnic University, [2]Aix Marseille Université, CNRS, Laboratoire Parole et Langage UMR 7309, 13100 Aix-en-Provence, France

ctyaoyao@polyu.edu.hk, Christine.Meunier@lpl-aix.fr

Phonological similarity has long been known as one of the most baffling lexical relation. Words that sound similar to each other (i.e. phonological neighbors) can be both friends and foes, depending on the nature of the processing task. In particular, it has been shown that similar-sounding words are competitors in perception, causing slower and less accurate word recognition (Luce and Pisoni, 1998), but they help each other in production tasks, as words with many neighbors are produced faster and more accurately than those with fewer neighbors (Vitevitch, 2002). Gahl *et al.* (2012) reported that in spontaneous speech, words from dense neighborhoods were more reduced phonetically than those from sparse neighborhoods, suggesting that micro-level phonetic variation in spontaneous production was more heavily influenced by speaker-oriented forces than listener-oriented forces.

However, most previous research on phonological neighborhood density was conducted on the English language, with only a few exceptions on Spanish and other languages. Furthermore, the controversy over the Spanish results (Baus *et al.*, 2008; Vitevitch and Stammer, 2006, 2009) brings into question whether neighborhood effects are consistent cross-linguistically. In this paper, we examine another Romance language, French, for the effects of neighborhood structure on phonetic variation. Previous literature has reported inhibitory effects of neighborhood density on French word recognition in (Dufour and Frauenfelder, 2012), though the effect on French word production is less clear (see Sadat *et al.* 2014 for a recent discussion).

Our data came from the CID corpus (Bertrand *et al.*, 2008), which consisted of eight hours of recordings of one-to-one dialogues from 16 speakers (10F). We extracted from the corpus a set of 5,412 tokens of monosyllabic monomorphemic content words between 1 and 5 phonemes long. Similar to Gahl *et al.*'s study, we constructed two statistical models on the variation of word duration and degree of vowel dispersion, respectively. The word duration model showed that neighborhood density (normalized for word length in phonemes) was a positive predictor for word duration ($\beta = 0.034$, $t = 2.12$, $p_{MCMC} = .03$), after controlling for a large number of lexical and contextual factors (baseline duration, lexical frequency, phonotactic probability, part of speech, orthographic length, phonemic length, contextual predictability, speech rate, previous mentions, speaker sex, etc.), but neighbor frequency failed to reach significance in the duration model ($t = 0.42$, $p_{MCMC} > .1$). That is to say, words with high neighborhood density tended to be produced with longer duration in French. General results of the duration model persisted when only 2- or 3-phoneme word tokens were tested. Similarly, the vowel dispersion model found that high-density words had significantly more dispersed vowels ($\beta = 0.12$, $t = 3.23$, $p_{MCMC} = .001$), when everything else was controlled (most of the control factors in the duration model plus of vowel type, vowel duration and phonetic context), with neighbor frequency being non-significant ($t = 1.51$, $p_{MCMC} > .1$) again. General findings of the vowel dispersion model persisted across a set of alternative models.

Taken together, the current modeling results suggest a pattern in French that is opposite of the one observed in English. Words from dense neighborhoods in the French lexicon are hyperarticulated - as opposed to hypoarticulated - compared with words from sparse neighborhoods. Is the discrepancy due to the inhibitory effects of neighborhood density on French word production (as suggested in Sadat *et al.*, 2014)? Or does it suggest language-specific patterns of phonetic variation, given the linguistic differences between English and French (in morphology, prosody, speech rhythm, etc.)? We discuss each possibility and the implications for the theories of lexical activation and phonetic variation.

References

Baus, C., Costa, A., & Carreiras, M. (2008). Neighbourhood density and frequency effects in speech production: A case for interactivity. *Language and Cognitive Processes*, 23(6), 866-888. / Bertrand, R., Blache, P., Espesser, R., Ferré, G., Meunier, C., Priego-Valverde, B., &Rauzy, S. (2008). Le CID: Corpus of Interactional Data. Annotationet Exploitation Multimodale de Parole Conversationnelle. *Traitement Automatique des Langues*, 49(3). / Dufour, S., & Frauenfelder, U. H. (2010). Phonological neighbourhood effects in French spoken-word recognition. *The Quarterly Journal of Experimental Psychology*, 63(2), 226-238. / Gahl, S., Yao, Y., & Johnson, K. (2012). Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of memory and language*, 66(4), 789-806. / Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and hearing*, 19(1), 1-36. / Sadat, J., Martin, C. D., Costa, A., & Alario, F. (2014). Reconciling phonological neighborhood effects in speech production through single trial analysis. *Cognitive Psychology*, 68, 33-58. / Vitevitch, M. S. (2002). The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(4), 735-747. / Vitevitch, M. S., & Stamer, M. K. (2006). The curious case of competition in Spanish speech production. *Language and Cognitive Processes*, 21(6), 760-770. / Vitevitch, M. S., & Stamer, M. K. (2009). The influence of neighborhood density (and neighborhood frequency) in Spanish speech production: A follow-up report, in Spoken, *Language Laboratory Technical Report*, 1, 1-6, University of Kansas.

[DAY 2: Oral Session 8]

Some cognitive factors behind vowel lengthening in spontaneous Japanese: A corpus-based study

Yasuharu Den
Faculty of Letters, Chiba University

Vowel length in continuous speech is determined by various factors. Previous studies in phonetic science and speech synthesis have identified the following factors, among others, that may affect the duration of a vowel: i) the inherent duration of the vowel, ii) the compensatory effect of the surrounding consonants, iii) the syntactic category of the word, iv) the position of the vowel in the word, the prosodic phrase, and the utterance, and v) the overall speech rate of the speaker. When we turn our attention to spontaneous discourse, however, we find other higher-level factors triggering the lengthening of vowels. Since in spontaneous discourse speakers need extra time to plan what to say next and to formulate it in an appropriate construction, they may suspend the ongoing utterance, stretching the speech segment being articulated, when they experience a severe cognitive load.

In this talk, I will investigate vowel lengthening in spontaneous Japanese, based on a quantitative analysis of a large-scale corpus. In particular, I focus on three locations at which vowel lengthening frequently occurs: the final vowel i) of an utterance preface (filler or discourse marker) such as *e* and *de*, ii) at the end of a preceding utterance, and iii) of a topic phrase marked by the particle *wa*. The following example illustrates these three sites:

… suru-n-desu-keredo-mo:	(1.1)	e:	(0.3)	saiaku-na-no-wa:	zieetai-ni …
do-N-POL-yet		um		worse-COP-N-TOP	SDF-DAT
Preceding utterance		Preface		Topic	

A series of our studies have shown that the durations of vowels at these locations are positively correlated with the length of the following utterance (*zieetai-ni …*) and that they are also affected by the strength of connectivity between the preceding and the current utterances. By using novel statistical modeling and incorporating the above mentioned lower-level factors into the models, I will demonstrate how vowel lengthening in spontaneous Japanese is governed by cognitive factors, more precisely, by planning of utterance and discourse.

[DAY 2: Poster Session 2]

Rendaku in Spontaneous Speech

Shin-Ichiro Sano
Okayama Prefectural University
shinichirosano@gmail.com

Background: Rendaku (sequential voicing) is a process where the initial consonant of the second member of morphologically derived words becomes voiced, as in /oo+tako/ => [oodako] "big octopus". Not all morphologically derived words undergo this voicing process, however, and there are many phonological and morphological conditions that affect the likelihood of rendaku application (e.g. Vance 1979; Itô and Mester, 1986). One famous example is Lyman's Law (or OCP[voice]), where rendaku is blocked when the second member of a compound already contains a voiced obstruent; e.g. /oo+tokage/ => [oo+tokage], *[oo+dokage] "big lizard".

Rendaku has long been subject to phonological studies with various approaches, such as OT (Itô and Mester, 2003) and other experimental approaches (Vance, 1979; Kawahara, 2012), and the phenomenon has contributed to many theoretical debates. In particular, judgment-based studies have recently been conducted, which have addressed the psychological reality of rendaku-affecting patterns in terms of *judgments*. However, the aspects of actual rendaku *production*, especially in spontaneous speech, are yet to be explored. With this background, this research 1) studies patterns of rendaku (non)-application in actual speech, 2) examines the restrictions that have been claimed to affect rendaku application proposed in the previous work, and 3) explores the potential effects of previously unknown factors, focusing on the production in spontaneous utterances. This corpus-based study offers new findings that could not have been obtained by more traditional, intuition-based and dictionary-based studies.

Method: Data were retrieved from the CSJ-RDB (*the Corpus of Spontaneous Japanese-Relational Database*, Kokuritsu-Kokugo-Kenkyuujo, 2012). The CSJ-RDB consists of 201 speech samples, amounting to 45 hours of speech. The data collection was conducted using Navicat Premium (http://www.navicat.jp/premium/), with reference to phonological and morphological information annotated to the CSJ-RDB. An exhaustive search for the data in the CSJ-RDB brought forth a total of 1,153 tokens of potential undergoer of rendaku. Whether a segment underwent rendaku or not was coded based on the annotation in the CSJ-RDB. Among those 1,153 words, 709 tokens showed rendaku (59.4%), and 484 did not. Each token was subjected to the analysis in terms of the factors that can affect rendaku application. All distributions below were shown to be statistically significant.

Results & Conclusion: Firstly, the following restrictions were shown to categorically block rendaku, in support of the previous proposals (e.g. Kindaichi, 1976): 1) The presence of voiced obstruent(s) in the second member (the effect of Lyman's Law); 2) the honorific prefix *o*-; and 3) some *kunyomi* readings (native readings) of numerals, such as *hito* 'one,' *huta* 'two'. **【Effect of lexical strata】**Although foreign words do not undergo rendaku, Sino-Japanese words, old borrowings from Chinese, can undergo rendaku (Yamato: 66.9%, Sino-Japanese: 44.7%), contra previous proposals. **【Effect of preceding consonants】** Gemination and rendaku never co-occur, as rendaku would produce voiced geminates, which are universally marked (e.g. ***ik-k**ai* 'first floor' => **ig-g**ai*). Also, syllabic nasals show higher probability of rendaku (63.1% vs. vowels: 57%). This can be due to the effect of *NC constraint that bans nasal-voiceless obstruent cluster (Pater, 1999). **【Effect of place】** The place of articulation produces the following hierarchy of rendaku likelihood that is consistent with the universal markedness hierarchy (*g >> *d >> *b): bilabial (82.8%) > alveolar (61.4%) > velar (54.4%). **【Effect of neighboring vowels】** Both preceding and following vowels show similar effects on rendaku: the probability of rendaku is higher next to back vowels (/a, o, u/), and lowest next to /e/. **【Effect of non-linguistic factors】**Rendaku was shown to be immune to non-linguistic, sociolinguistic factors (e.g. style), except for gender. This implies that rendaku is a purely linguistic process, and is not consciously controlled.

In short, in spontaneous speech, the application of rendaku is controlled by various kinds of phonological/morphological factors. The results are cross-linguistically relevant, as some of the factors reflect markedness constraints that are independently motivated in other languages (e.g. universal markedness hierarchy on voiced stops, *VoicedGeminate, and *NC). In summary, rendaku tends to avoid universally marked structures.

References

Itô, J. and A. Mester. 1986. The phonology of voicing in Japanese: Theoretical consequences for morphological accessibility. *Linguistic Inquiry* 17, 49-73. / **Itô, J. and A. Mester. 2003.** *Japanese Morphophonemics*. Cambridge: MIT Press. / **Kawahara, S. 2012.** Lyman's Law is active in loanwords and nonce words: Evidence from naturalness judgment experiments. *Lingua* 122(11), 1193-1206. / **Pater, Joe. 1999.** Austronesian nasal substitution and other NC effects. In: Kager, R., H. van der Hulst, and W. Zonneveld (eds.), *The Prosody-Morphology Interface*. Cambridge: Cambridge University Press, 310-343. **Vance, T. 1979.** *Nonsense word experiments in phonology and their application to rendaku in Japanese*. Doctoral dissertation, University of Chicago.

Catalan prepalatal allophony and phonological contrast

José I. Hualde[1], Christopher Eager[1], Marianna Nadeu[2]
[1]University of Illinois at Urbana-Champaign, [2]Penn State
jihualde@illinois.edu, eager2@illinois.edu, nadeu@psu.edu

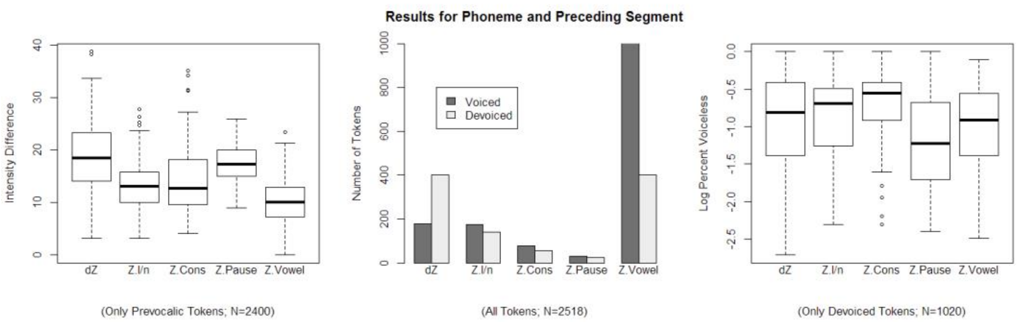
Classical Latin did not have palatal or prepalatal consonants. These segments arose in Late Latin from a number of sound changes. In particular, a prepalatal voiced affricate /ʤ/ developed primarily from the fortition of syllable-initial glides (IANUARIU > Italian *gennaio* 'January') and from the palatalization of velars and dentals (DIURNALE > It. *giornale* 'daily'). This consonant remains an affricate in Italian, but was later systematically lenited to a fricative /ʒ/ in French and Portuguese. In Central Catalan (Cat), both [ʒ] and [ʤ] are found. Word-initially, fricative and affricate realizations are, for the most part, in complementary distribution, depending on the preceding context, although a certain degree of free variation is sometimes described (Recasens, 1993:160, Wheeler, 2005:14). In a pilot study with read speech (N = 239) we find variation between fricative and affricate realizations of the phoneme /ʒ/ in all postconsonantal contexts. Word-medially and word-finally, however, fricative and affricate are different phonemes. That is, between vowels, /ʒ/ and /dʒ/ are in phonological contrast if the consonant is word-medial or word-final, but not if it is word-initial. Another aspect of variation has to do with voicing. Word-finally, voiced and voiceless obstruents are said to be in complementary distribution, conditioned by the following context in the phrase, and, additionally, word-medial (intervocalic) /dʒ/ may be devoiced in certain words or contexts. As far as we know, there are no previous quantitative, acoustic studies of this complex allophony in Cat. This study is of interest both because of the complexity of the phonological facts and because of the light that it may potentially shed on sound change.

To investigate the factors that contribute to the allophony of Cat /ʒ/ and /dʒ/, we conducted a large-scale corpus study with data from *Glissando*, a corpus of Cat and Spanish speech (Garrido et al., 2013), containing both scripted and unscripted speech from 28 speakers. To examine the allophony of these consonants, we focus on an acoustic correlate of degree of constriction, intensity difference (IntDiff), i.e., the difference between the lowest intensity value in the consonant and the maximum in the following vowel. A larger IntDiff value indicates a more constricted realization (see, e.g., Hualde et al., 2011). The data were analyzed with linear mixed effects regression. In addition to the expected difference between /ʒ/ and /ʤ/, we find an effect of the preceding context (after pause > after n/l > after another C > after V), in line with the traditional view. At the same time, the fact that postconsonantal tokens are not as constricted as postpausal ones, together with the degree of overlap that we find among contexts (Fig. 1a), would argue for something more complex than two allophones, fricative and affricate, in complementary distribution. We also find a significant effect of stress and style (scripted > unscripted). There was, however, no effect of word-position. Other things being equal, word-medial /ʒ/, which contrasts with /ʤ/, does not differ in constriction from noncontrastive word-initial /ʒ/.

Percentage of devoicing was also calculated for each token. 69% percent of all tokens of /ʤ/ and 33% of tokens of /ʒ/ were found to be devoiced to some degree. Partial devoicing is thus common in all contexts, especially for the phonological affricate (Fig 1b). Full devoicing, on the other hand, is very rare in our corpus (1% voiceless tokens of /ʤ/ and 0.5% of /ʒ/). In word-final position, the voice contrast is neutralized in Cat. We considered only tokens before a vowel or a voiced consonant, where assimilation in voice has been described as being categorical. The amount of devoicing in word-final position is not significantly greater than in other contexts.

An important finding, regarding both IntDiff and voicing, is that the allophony of the target consonants appears to be conditioned purely by phonetic factors and not by the existence or absence of a phonological contrast (in manner or voicing) in a given position in the word.

Figure 1: a) IntDiff by preceding context; b) Percent of partially devoiced tokens by context; c) Degree of devoicing



References

Garrido, J. M. et al. 2013. Glissando: a corpus for multidisciplinary prosodic studies in Spanish and Catalan. *Lang. Resour. Eval.* 47, 945-971. / Hualde, J. I. et al. 2011. Consonant lenition and phonological recategorization. *Lab. Phon.* 2, 301-329. / Recasens, D. 1993. *Fonètica i fonologia*. Barcelona: Enciclopèdia Catalana. / Wheeler, M. 2005. *The phonology of Catalan*. Oxford: OUP.

Revisiting vowel harmony in Korean sound-symbolic words: a corpus study

Sang-Im Lee-Kim
New York University
sangim@nyu.edu

Korean vowel harmony refers to a harmony pattern involving a set of *light* [a, o] and *dark* vowels [i, ɨ, u, e, ʌ] (light harmony: [a]l.l[o]k.t[a]l.l[o]k 'colorful' *vs*. dark harmony: t[u]ng.k[ɨ]l.t[u]ng.k[ɨ]l 'round') within sound-symbolic adjective roots or between verb-suffix pairs. Studies on harmony in sound-symbolic words have shown that the majority of the attested forms conforms to the harmony rule, but a small subset of roots allow disharmonic forms as well (Hong 2010, Larson and Heinz 2012, L&H hereafter). While neutral vowels in disharmonic roots are traditionally known to be [i, ɨ] in non-initial positions (e.g. p[a]ng.k[ɨ]s.p[a]ng.k[ɨ]s 'smiling'), L&H argue that [u, y] in addition are at least partially neutral to vowel harmony. The present study carried out a large-scale corpus study to verify the status of the neutral vowels in Modern South Korean (MSK) sound-symbolic words. The results showed that neutral vowels in Korean include [i, ɨ, u] in line with L&H, but only [ɨ] is truly transparent in medial positions (e.g. p[a].s[ɨ].l[a]k 'rustling'), whereas other neutral vowels are all opaque. We interpret the results as an indication that the vowel with the most unmarked status in MSK phonology is [ɨ].

While L&H seeks to capture the synchronic grammar of Korean, there are two problematic concerns regarding this data. First, they used a sound-symbolic dictionary (SSD) rather than a real corpus, raising concerns about the reliability of the data. In particular, the specific dictionary has been subject to criticisms that it lists many obsolete forms (1a) and those exclusively used in North Korea (1b) which are not intelligible to Modern Korean speakers.

(1) Examples of disharmonic roots used in previous corpus studies

a. non-MSK words		b. North Korean words	
t[a].p[u]l.t[a].p[u]l	'flowing hair of a kid'	k[a].k[u]l.k[a].k[u]l	'windingly'
p[o]ng.s[i]s.p[o]ng.s[i]s	'a light smile'	n[o].c[i]l.n[o].c[i]l	'warbling of a lark'

Second, while L&H assume a 10-vowel system including [y, ø], most scholars no longer posit those vowels as contemporary monophthongs in MSK. This undermines L&H's conclusion that [y] is neutral to vowel harmony in Modern Korean. While it is common to assume an 8-vowel system, we posit a 7-vowel system to reflect the most recent MSK phonology, considering nearly complete merger of [ɛ] to [e].

In place of a dictionary, the current study used the 'Sejong corpus (SJ)' which consists of 10 million tokens (90% of written materials from newspapers, magazines, novels/essays, etc., and 10% of spoken materials from conversations in everyday life, movie subtitles, TV shows, plays, etc.). To construct a data set that only contains the existing sound-symbolic words, an intersection of SSD and SJ was drawn using a python script. If a sound-symbolic word in SSD is found in SJ, it is considered to be an existing word in MSK, whereas if not, it is discarded from the dataset. Our new corpus consisted of 927 sound-symbolic words in total (733 disyllabic and 194 trisyllabic words), dramatically reduced from about 4,000 words reported in previous studies. However, the ratio of the disharmonic roots (74/1,068= 6.9%) actually increased as compared to L&H (133/3,972 = 3.3%).

To test the status of the arguably neutral vowels, uncontroversial dark vowels [e, ʌ] are used as references for a statistical comparison following L&H. The results were in accordance with L&H that i) [u, i, ɨ] are dark in initial, but neutral in non-initial positions and that ii) the neutrality of [u] is weaker than [i, ɨ]. However, the results dispute L&H as well as the traditional view on the transparent vowels: only [ɨ] is transparent (✓ L-ɨ-L), but [u, i] are largely opaque (*L-u/i-L) in medial positions.

The present study points to the important role of a corpus study in verifying theoretical claims. We however show that the validity of the corpus is essential to the reliability of the results. Moreover, the corpus study sheds light on the special status of the vowel [ɨ] in MSK. That is, the rather striking result of [ɨ] being the only transparent vowel in MSK suggests that it is the least marked vowel in MSK phonology. Supporting evidence comes from independently motivated phonological processes: [ɨ] is used as an epenthetic vowel when breaking up consonant clusters in loanwords (e.g. /ski/ -> [sɨ.ki] 'ski') (Oh, 1992), and reversely is most susceptible to coalescence or deletion (e.g. /ɨi/ → [e] 'poss. marker') (Sohn, 1987). Phonetically, [ɨ] is subject to strong coarticulation with preceding consonants (Kim and Kochetov, 2011), providing a phonetic source of the vowel being transparent to vowel harmony.

References

Larson, D., Heinz, J. 2012. Neutral vowels in sound-symbolic vowel harmony in Korean. *Phonology* 29, 433-464. / Hong, S. H. 2010. Gradient vowel cooccurrence restrictions in monomorphemic native Korean roots. *Studies in Phonet., Phonol. and Morph.* 16. 279–295. / Kim, K., Kochetov, A. 2011. Phonology and phonetics of epenthetic vowels in loanwords. *Lingua* 121, 511–532. / Oh, M. 1992. Palatal consonants, labial consonants, and vowel epenthesis in Korean loanword phonology. *Ling. J. of Korean*, 17, 141–161. / Sohn, H. S. 1987. On the representation of vowels and diphthongs and their merger in Korean. *Chicago Ling. Soc.* 23, 307–323.

Advancing corpus-based analyses of spontaneous speech: Switch to GECO!

Antje Schweitzer, Natalie Lewandowski & Grzegorz Dogil
Institute for Natural Language Processing, Stuttgart University
antje.schweitzer@ims.uni-stuttgart.de

We present the GECO (**GE**rman **CO**nversations) database, which was designed for investigating phonetic convergence in German and its correlation with personality traits and scores of mutual likeability and esteem. The data can more generally be used to investigate aspects of interlocutors' behavior in free, spontaneous dialogs, including social aspects. The database will be freely available for non-commercial use and to our knowledge will be by far the largest German database of this type. GECO consists of 46 dialogs of approx. 25 minutes length each, between previously unacquainted female subjects. It further contains results of personality tests aimed at subjects' extraversion, acting behavior, other-directedness and sensitivity to social cues (all collected once per subject), as well as participants' mutual ratings in terms of competence and likeability (collected after every dialog). 22 dialogs took place in a unimodal (UM) setting, where participants were separated by a solid wall and could not see each other, while the remaining 24 dialogs were recorded with subjects facing each other (multimodal setting, MM). All dialogs were recorded in high quality (separate channels, 16 bit@48 kHz) in a sound-attenuated booth using AKG HSC271 headsets with rubberfoam windshields. In the MM setting, a transparent screen ensured sufficient speaker separation between the two channels. We had 12 speakers in the UM condition, and 8 (7 from UM, 1 additional speaker) in the MM condition, which was recorded 5 months later. All subjects were females between 20 and 30 years of age, mostly students. They were paid for each dialog they participated in. Subjects were naïve to the research questions; they were told that the purpose of the study was to research how small talk between strangers works. They were provided with a list of potential topics to ease conversation, but were explicitly told that they were completely free to choose other topics as well.

Most recent studies on phonetic convergence and imitation use rather controlled and limited speech material, often without real conversational interaction, or focus on only specific target words or phrases in conversations (Delvaux & Soquet 2007, Babel 2010, Abrego-Collier et al. 2011, Kim et al. 2011, Nielsen 2011, Babel 2012, Pardo et al. 2012). Few recent studies use larger-scale fully annotated corpora such as the quasi-spontaneous Columbia Games Corpus (Levitan & Hirschberg 2011), which does not provide social or personality factors. However, social factors are assumed to be central in convergence. The GECO database was conceived to close this gap, providing social data in addition to large-scale fully annotated recordings of high audio quality as a basis for corpus-based approaches to phonetic convergence or other aspects of conversational behavior.

The GECO database was automatically processed using a number of tools: Starting from manual annotations, which included hesitations, filled pauses, and restarts, we automatically generated expected pronunciations using the Festival speech synthesis system (www.festvox.org) including German CELEX (Baayen et al. 1995) and an in-house morphology component to alleviate the high number of out-of-vocabulary words, which is due to the high morphological productivity of German. The spontaneous nature of our data called for the prediction of pronunciation variants in addition to the canonical pronunciations generated by our system. We then annotated all data on the segment, syllable and word level using forced alignment (Rapp, 1995), letting the alignment tool decide where variants were used instead of canonical forms. We then parameterized the F0 contours and automatically generated prosodic annotations (German GToBI labels) using classifiers trained on read data. Preliminary results indicate good precision but low recall. The quality of the automatically generated labels, which is of course not comparable to that of manual annotations, is subjectively good enough to be valuable as additional information to complement continuous prosodic parameters for instance. The resulting corpus amounts to 20.7 hrs. of dialog (two channels), with approx. 250,000 words, 360,000 syllables, 870,000 phones, 46,000 pitch accents, and 28,000 phrase boundaries. For illustration we will present experimental results (convergence of vowel formants, articulation rate, and backchannels) based on this corpus.

References

C. Abrego-Collier, J. Grove, M. Sonderegger, & A.C. L. Yu, 2011. Effects of speaker evaluation on phonetic convergence. *Proc. 17th ICPhS Hong Kong*, 192-195. / H. Baayen, R. Piepenbrock, & L. Gulikers, 1995. *The CELEX lexical database - Release 2*, Centre for Lexical Information, Max Planck Institute for Psycholinguistics, Nijmegen. / M. Babel, 2010. Dialect divergence and convergence in New Zealand English. *Language in Society* 39(4), 437-456. / M. Babel, 2012. Evidence for phonetic and social selectivity in spontaneous phonetic imitation, *J. Phon.* 40(1), 177-189. / V. Delvaux & A. Soquet, 2007. The influence of ambient speech on adult speech productions through unintentional imitation, *Phonetica* 64, 145-173. / M. Kim, W. S. Horton, & A. R. Bradlow, 2011. Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *J. Lab. Phon* 2, 125-156. / R. Levitan & J. Hirschberg, 2011. Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. *Proc. 12th Interspeech Florence*, 3081-3084. / K. Nielsen, 2011. Specificity and abstractness of VOT imitation. *J. Phon.* 39(2), 132-142. / J. S. Pardo, R. Gibbons, A. Suppes, & R. M. Krauss, 2012. Phonetic convergence in college roommates. *J. Phon.* 40(1), 190-197. / S. Rapp, 1995. Automatic phonemic transcription and linguistic annotation from known text with Hidden Markov models - An aligner for German. In: *Proceedings of ELSNET Goes East and IMACS Workshop "Integration of Language and Speech in Academia and Industry"*, (Moscow, Russia).

Phonetic and Prosodic Characteristics of Disfluencies in French Spontaneous Speech

George Christodoulides[1], Mathieu Avanzi[2]

[1] University of Louvain, Belgium, [2] University Paris Diderot, France

george@mycontent.gr, mathieu.avanzi@gmail.com

A key difference between **spontaneous speech** and controlled laboratory speech is the prevalence of **disfluencies** in the former (e.g. Shriberg 1994). Disfluencies typically signal production problems, as the speaker incrementally constructs his message (Levelt 1989). However, in specific contexts, these events may be used as communicative devices, e.g. in order to manage dialogue interaction (Moniz et al. 2009) or indicate information status (Arnold et al. 2003). Disfluencies have recently attracted the interest of the phonetic sciences and of computational linguists working on speech (e.g. Adda-Decker et al. 2004, Arbisi-Kelm 2010, Germesin et al. 2008, Stolcke & Shriberg 1996).

The **corpus** used in this study consists of 14 regional varieties of French, recorded in France, Switzerland and Belgium, with 112 native speakers (4 male and 4 female speakers per regional variety, aged between 20 and 80). We focus on semi-directed interviews, which are conversations where the interviewee produces many continuous stretches of speech and has little interaction with the interviewer. The corpus has been orthographically transcribed, aligned to the phone and syllable level, and POS tagged. The corpus was annotated under Praat (Boersma & Weenink 2014).

We propose a detailed annotation scheme for disfluencies, based on the typology of Shriberg (2001) and Brugos et al. (2012). **Simple** disfluencies include filled pauses (autonomous fillers, i.e. surrounded by silent pauses, as well as epenthetic vowels), mispronunciations, segment lengthening (drawls), and inter-word pauses. **Structured** disfluencies follow the pattern "reparandum – interruption point – optional editing terms – repair" and include repetitions, insertions, substitutions and deletions. **Complex** disfluencies (i.e. sequences of several simple and/or structured disfluencies) are annotated according to Heeman et al. (2006), leading to a fine-grained description of such events.

We developed software to facilitate the annotation and the extraction of **phonetic and prosodic features** of disfluencies and their context. The features under study are duration, pitch (average, register and contour), pitch and energy slopes, voice quality and vowel quality (F1 and F2 formants of fillers). We also take into consideration speaking rate and the morphosyntactic categories of disfluent tokens, as well as their left and right context.

Our findings are in line with previous studies on the phonetic and prosodic features of disfluencies (e.g. for English, Shriberg 1999, 2001; for French, Vasilescu et al. 2004; for Portuguese, Moniz et al. 2012). In the **reparandum** region of structured disfluencies, the syllables affected are those immediately preceding the interruption point (lengthening or trailing fillers or short silent pauses). The **vowel length distribution** of fillers is significantly different from the distribution of vowel length in fluent contexts, distinguishing disfluency-related lengthening from stress-related prominence. **Filled pauses** exhibit falling intonation contours, with a mean pitch close the speaker's mean pitch; however the formant values of vowels in filled pauses are significantly different from those of the same vowels in fluent contexts. Finally, pitch and energy increase in the **repair** region (compared to the reparandum) possibly signalling the boundaries of each region. We are currently exploring to what extent these observations are affected by **regional variation** and more detailed results will be presented at the conference.

References

Adda-Decker, M., Habert, B., Barras, C., Boula de Mareüil, Ph., Paroubek, P. 2004. Une étude des disfluences pour la transcription automatique de la parole et l'amélioration des modèles de langage. *Actes des 25èmes JEP*. / Arbisi-Kelm, T. 2010. Intonation structure and disfluency detection in stuttering, In: Fougeron, C., Kühnert, B., D'Imperio, M., Vallé, N. (eds.) *Laboratory Phonology* 10. Berlin: De Gruyter Mouton, 405-432. / Arnold J. E., Fagnano M., Tanenhaus M. K. 2003. Disfluencies signal theee, um, new information, J. Psycholinguist Res, 32(1), 25-36 / Boersma, P., Weenink, D. 2014. Praat: doing phonetics by computer, ver. 5.3.63, www.praat.org / Brugos, A., Shattuck-Hufnagel, S. 2012. A proposal for labelling prosodic disfluencies in ToBI. *Poster, Advancing Prosodic Transcription for Spoken Language Science and Technology*, July 31, 2012, Stuttgart, Germany. / Germesin, S., Becker, T., Poller P. 2008. Hybrid Multi-step Disfluency Detection, *Proc. 5th Intl Workshop on Machine Learning for Multimodal Interaction*. / Heeman, P., Mcmillin, A., Scott Yaruss, J. 2006. An annotation scheme for complex disfluencies. *Proc. 9th ICSLP*, 1081-1084. / Levelt, W. J. M. 1989. *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press / Moniz, H., Trancoso, I., Mata da Silva, A.I., 2009. Classification of disfluent phenomena as fluent communicative devices in specific prosodic contexts, *Proc. Interspeech 2009, Brighton, UK*, 1719-1722. / Moniz, H., Batista, F., Trancoso, I., Mata da Silva, A.I. 2012. Prosodic context-based analysis of disfluencies. *Proc. Interspeech Portland, Oregon*. / Shriberg, E. 1994. Preliminaries to a Theory of Speech Disfluencies, Ph.D. thesis, University of California, Berkeley. / Shriberg, E. 1999. Phonetic Consequences of Speech Disfluency. *Proc. 14th ICPhS San Francisco*, 619-622. / Shriberg, E. 2001. To 'errrr' is human: ecology and acoustics of speech disfluencies, *J. Intl. Phonetic Assoc.* 31(1), 153-169. / Stolcke, A., Shriberg, E. 1996. Statistical language modeling for speech disfluencies. *Proc. ICASSP96*, 405-408. / Vasilescu, I., Candea, M., Adda-Decker, M. 2004. Hésitations autonomes dans 8 langues : une étude acoustique et perceptive. *Actes de MIDL 2004*, 25-30.

Multilevel Modelling of Initial Rise in French

Mathieu Avanzi[1], *George Christodoulides*[2]
[1]University Paris Diderot, France, [2] University of Louvain, Belgium
mathieu.avanzi@gmail.com, george@mycontent.gr

In French, the minimal prosodic unit is the Accentual Phrase (AP). An AP is tonally marked by a high pitch movement associated with its rightmost full syllable and an optional high tone aligned with a syllable on its left (cf. the default pattern LHiLH*, identified by Jun & Fougeron 2002). These two tonal events are accompanied by durational cues: the final rise is marked by pre-boundary lengthening; and the initial rise is often marked by consonant onset lengthening. There is a general agreement in the literature regarding the fact that final rises are pitch accents, and the interactions of the numerous factors that influence their distribution are well-studied. Conversely, the phonological status of initial rises is still a question open to debate (Welby 2006), and the factors favouring its realisation are more poorly understood (Di Cristo 1998).

Portes et al. (2012) recall that the initial rise (IR) does not have a fixed position: it can appear on a function word or on one of the first syllables of the content word of the AP, the first syllable of this content word being the most common position (Astésano et al. 2007). It has been claimed that the optional realisation of initial rises depends on metrical structure (the more syllables the AP contains, the greater chances for an IR to be realised), on speaking rate (the faster the speaker articulates, the greater the chances for an IR not to be realised), or on the position of the AP within the host intonational phrase (the IR would preferentially appear at the edge of pre-nuclear APs rather than at the edge of nuclear APs). It has also been claimed that speaking style has an effect on IR distribution: formal situations (such as political discourse, didactic speech, broadcast speech) and read-aloud speech would favour the production of initial rises (Lucci 1983). In addition, it has been shown that the geographic origin of the speaker is an important factor affecting the realisation of initial rises: Boula-de-Mareüil et al. (2012) show that speakers originating from the French-speaking part of Switzerland tend to realise more initial rises than Belgian of Hexagonal French speakers.

To our knowledge, the way these constraints interact to define initial rise distribution in French remains unknown. The aim of this paper is to provide a multilevel modelling of initial rise realisation in spontaneous French, by evaluating the weight and the interactions of the factors mentioned above, based on corpus evidence. For this purpose, a 12-hour-long annotated corpus was used. The corpus consists of 14 regional varieties of French recorded in France, Belgium and Switzerland, with 112 native speakers (4 male and 4 female speakers per regional variety). Each speaker was recorded in a reading text task and a conversation task. The entire reading text and a stretch of 3 minutes of spontaneous speech for each speaker was orthographically transcribed and automatically aligned with the EasyAlign script within Praat (Boersma & Weenink 2014). The alignment was manually checked and corrected. Initial rises and pitch accents were identified independently by two experts, on the basis of their perceptual judgment only. AP boundaries (i.e. Clitic Groups carrying a pitch accent on their rightmost syllable) and intonation phrase boundaries (pitch accented syllables followed by a pause, associated with a major pitch rise and/or significant lengthening) were automatically identified and annotated on separate tiers.

Results reveal that there is a significant effect of the grammatical status of the word (lexical words attract initial rise to a greater extent than function words), an interaction between the position of the syllable within the word and the number of syllables of the word (initial rises are preferentially associated with the first syllables of dissyllabic words), an effect of articulation rate (the mean shorter the mean syllabic duration of the host IP, the greater chances for IR to be realized), an effect of speaking style (a greater percentage of realised initial rises in the read speech sub-corpus than in conversation) and an effect of speakers' origin: speakers from Switzerland realise more initial rises than those from France and Belgium. Surprisingly, no effect of the position of the AP within the host IP (pre-nuclear vs. nuclear) was found. These results and their impact on phonological status of the Initial Rise will be further discussed during the conference.

References

Astésano, C., Bard, E. G., Turk, A. 2007. Structural Influences on Initial Accent Placement in French. *Language and Speech* 50, 423-446. / Boersma, P., Weenink, D. 2014. Praat: doing phonetics by computer, ver. 5.3.63, www.praat.org / Boula de Mareüil, P., Adda-Decker, M., Woehrling, C., Bardiaux, A., Simon, A. C. 2012. Une étude par le traitement automatique de la prosodie du français à la frontière des domaines romans et germaniques. In: A. C. Simon (ed.), *La variation prosodique régionale en français*. Bruxelles: De Boeck/Duculot, 119-136. / Di Cristo, A. 1998. Intonation in French. In: D. Hirst & A. Di Cristo (eds.), Intonation Systems: A Survey of Twenty Languages, Cambridge: Cambridge University Press, 195-218. / Jun, S. A., Fougeron, C. 2002. Realizations of Accentual Phrase in French Intonation. *Probus*, 14, 147-172. / Lucci, V. 1983. *Etude phonétique du français contemporain à travers la variation situationnelle*. Grenoble: Publications de l'Université de Grenoble. / Portes, C., D'Imperio, M., Lancia, L. 2012. Positional constraints on the initial rise in French. *Proc. 6th Speech Prosody*, 563-566. / Welby, P. 2006. French Intonational Structure: Evidence from Tonal Alignment. *Journal of Phonetics* 34, 343-371.

Stress-induced sibilant variations in Taiwan Mandarin spontaneous speech

Yu-Ying Chuang & Janice Fon

Graduate Institute of Linguistics, National Taiwan University

{d98142003; jfon}@ntu.edu.tw

The manifestation of prosody is not restricted to the suprasegmental domain. Instead, more and more evidence has surfaced, demonstrating that prosodic structure is also encoded in the fine-grained segmental details. In specific relation to stress, a greater articulatory effort is usually devoted to the prosodically prominent position (Beckman et al., 1994). However, the strengthening details differ both across and within languages. For example, de Jong and Zawaydeh (2002; 2004) found that discrepant vowel features are enhanced in Arabic and English. In addition, although prominence usually results in an enlarged phonemic contrast, Cole et al. (2007), on the contrary, reported that the acoustic cues defining the place of articulation contrast are uniformly enhanced toward the same direction, which is termed as *uniform strengthening*. It could therefore be inferred from these studies that the stress effect is actively interacting with various linguistic factors. To further explore this issue, it is the goal of this study to investigate the stress effect in a sociolinguistic context. Specifically, both the prominent and reduced stress levels are looked into, and we are interested in whether stress is diversely encoded in the segmental signals by different social speaker groups.

The target segments are dental and retroflex sibilants in Taiwan Mandarin. The two sibilants were chosen because they are sociolinguistically-sensitive phoneme pairs in Taiwan Mandarin. As indicated by previous studies, there exists a stigmatized merging rule of substituting dental for retroflex sibilants (Kubler, 1985). Overall, females retain more retroflex productions than males, and northerners, who speak the standard variety, are generally more inclined to retroflex productions than southerners (Lin, 1983). It is thus expected that with the divergent baseline sibilant productions of different speaker groups, their stress strategies for realizing dental and retroflex sibilants should not be the same. Given the fact that the merging production is usually consciously avoided in the reading tasks due to its negative connotation, the present study examined the sibilant realizations in spontaneous speech.

There were three to four speakers for each speaker group (Northern Male/Northern Female/Southern Male/Southern Female), and 30-miniute-long interview speech was elicited from each participant. The stress labeling followed the criteria of the Pan-Mandarin ToBI (Peng et al., 2007), in which three non-lexical stress levels are distinguished: s3 (stressed), s2 (unstressed), and s1 (reduced). The inter-labeler reliability for the stress labeling was satisfactory. All dental and retroflex sibilants were identified, and their realizations were auditorily determined. The acoustic measurement for the sibilants was centroid frequency, which positively correlates with the advancement of the tongue position. Considering that centroid frequency is highly dependent on the length of the vocal tract, for comparisons, we paired up s3 and s1 sibilants with the s2 sibilants that were nearest to them within the same intonational phrase for each speaker, since the number of s2 sibilant tokens greatly outnumbered those of s3 and s1 ones.

Results indicated that for dental sibilants, s3 dentals were realized with higher centroid frequency than their s2 counterparts, and s1 dentals were realized with lower centroid frequency than their s2 counterparts. This forms a consistent trend, in which the place of articulation is more anterior as the stress level increases. The four speaker groups, however, did not exhibit discrepant patterns. For retroflex sibilants, the stress effect is similarly exerted; that is, the tongue position was more advanced for s3 retroflexes and less advanced for s1 retroflexes. In addition, significant speaker group differences were reported. Among northern speakers, the difference between s3 and s2 was made larger by males than females. Moreover, the s1-s2 distinction was greater for southern females than for northern females. Our results, comparable with Cole et al. (2007), suggested a uniform strengthening effect on sibilant realizations. With respect to the sociolinguistic influences, the lack of variation for dental sibilants mirrored the fact that dental sibilants were more stably produced than retroflex sibilants in Taiwan Mandarin. By contrast, for retroflex sibilants, it was observed that the more dental-prone group (males, southerners) would make a greater advancement of the tongue position to signal stress differences. To conclude, the present study has demonstrated how the stress effect on segmental realizations was influenced by sociolinguistic factors. A uniform strengthening/weakening effect on the tongue position was found for both dental and retroflex sibilants in Taiwan Mandarin. The extent to which the stress contrast was made, however, was closely related to the sibilant production of each speaker group.

References

Beckman, M. E., & Edwards, J. (1994). Articulatory evidence for differentiating stress categories. In P. A. Keating (Ed.), *Papers in Laboratory Phonology III: Phonological structure and phonetic form* (pp. 7–33). / Cole, J., Kim, H., Choi, H., & Hasegawa-Johnson, M. (2007). Prosodic effects on acoustic cues to stop voicing and place of articulation: Evidence from Radio News speech. *J. Phon.*, *35*(2), 180–209. / de Jong, K., & Zawaydeh, B. (2002). Comparing stress, lexical focus, and segmental focus: patterns of variation in Arabic vowel duration. *J. Phon.*, *30*(1), 53–75. / de Jong, K., & Zawaydeh, B. (2004). Stress, lexical focus, and segmental focus in English: patterns of variation in vowel duration. *J. Phon., 32*(4), 493–516. / Kubler, C. C. (1985). The influence of Southern Min on the Mandarin of Taiwan. *Anthropological Linguistics*, *27*, 156–176. / Lin, C. C. (1983). A sociolinguistic study of the use of the retroflex sounds in Mandarin in college studennts in Taiwan. *Bulletin, College of Arts and Letters, National Central University*, 1–15. / Peng, S., Chan, M. K. M., Tseng, C., Huang, T., Lee, O. J., & Beckman, M. E. (2007). Toward a Pan-Mandarin system for prosodic transcription. In S.-A. Jun (Ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press.

(t,d) Deletion in Everyday Speech

Margaret E. L. Renwick,[1] Rosalind Temple,[2] Ladan Baghai-Ravary,[2] and John S. Coleman[2]

[1]Program in Linguistics, University of Georgia, USA; [2]Phonetics Laboratory, University of Oxford, UK

mrenwick@uga.edu, rosalind.temple@mod-langs.ox.ac.uk,
{ladan.baghai-ravary, john.coleman}@phon.ox.ac.uk

We study variation in word-final /t/ and /d/ in consonant clusters, using the Audio British National Corpus (Coleman et al. 2012). This sociolinguistic variable has been extensively studied, and modelled in terms of Lexical Phonology: Guy (1991) argues that a word's morphological structure affects the likelihood of application of a variable deletion rule, such that deletion is most likely in monomorphemes (e.g. *accept*), less likely in irregular past tense forms (e.g. *kept*), and least likely in regular past tense forms (e.g. *stopped*). However, some corpus studies have failed to find a significant morphological effect (e.g. Tagliamonte & Temple 2005), or suggest that (t,d) variation is a continuous phonetic process (Temple in press), or find that lexical frequency accounts for much of morphology's purported effect (Guy et al. 2008).

We use the uniquely large and informal Audio BNC to examine the effects of morphology and lexical frequency while controlling for phonological context, studying /t,d/ as segments subject to more or less phonetic reduction or deletion. The data are word pairs in which /t,d/ appears in a consonant cluster at the end of the first word, in an array of phonological contexts; we compare 12 (near-)minimal triplets of a monomorpheme, an irregular past, and a regular past. To control phonological context we matched the pre-/t,d/ vowel and consonant as closely as possible; e.g. the /st#ɪ/ context compared *past eleven* (32 tokens), *lost in* (21 tokens), and *dressed in* (14 tokens). In addition to the triplets, in which the presence of /t,d/ may vary, for each context we collected two controls: a word pair in which /t,d/ is *always* present (*across to*, 32 tokens) and one which *never* contains /t,d/ (*unless it's*, 30 tokens). 5191 tokens (2037 test tokens) were collected across 63 word pairs. No judgment was made on the presence or absence of /t,d/. To detect phonetic reduction we measured intensity (dB) every 5ms throughout each pair; if /t,d/ is reduced or deleted, higher intensity levels are predicted just before the word boundary than when /t,d/ is a fully-occluded stop. Intensity was z-transformed (σ) in each token to allow comparison across tokens and contexts, and was temporally aligned at the word boundary across all tokens of each pair (Figure 1). To each token we fitted a third-order orthogonal polynomial, whose coefficients describe shape features of the intensity curve over the 100ms before the word boundary.

Results. We combined morphology, phonological context, and lexical frequency in a linear model of intercept coefficients from each polynomial. Intensity is lower in voiceless clusters (e.g. /st#ɪ/) than in voiced ones (e.g. /ld#ɪ/). Once voicing is accounted for, we find a significant correlation between intercept intensity and lexical frequency: most of the variation in mean intercept intensity is explained by the log probability of the first word in the pair (log P(W1)), and the voicing of its word-final consonant (Figure 2). When both log P(W1) and morphological category are included in the same model, log P(W1) is highly significant ($p \approx 0$) while morphology is not ($p > 0.5$). Additionally, we find a strong inverse correlation between the duration of W1 and its log probability. In sum: more frequent words are shorter and higher in intensity than less frequent words, which tend to be longer and end with lower intensity. We interpret this to mean that lower-frequency words are pronounced more slowly and carefully, leading to a lower incidence of (t,d) "deletion." Our results provide little support for the role of morphology as predicted by Lexical Phonology (Guy 1991).

Figure 1: Intensity over time across morphological categories

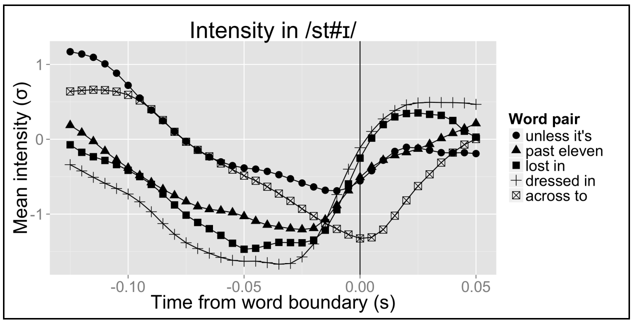
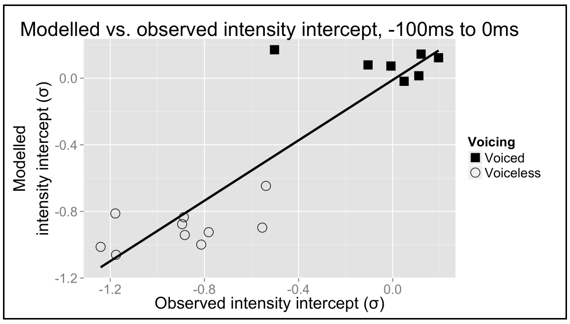


Figure 2: Intensity = f(log P(W1), voicing). R^2= 0.79



References

Coleman, John, Ladan Baghai-Ravary, John Pybus & Sergio Grau. 2012. *Audio BNC: the audio edition of the Spoken British National Corpus*. http://www.phon.ox.ac.uk/AudioBNC. / Guy, Gregory R. 1991. Explanation in variable phonology: An exponential model of morphological constraints. *Language Variation and Change* 3, 1–22. / Guy, Gregory R., Jennifer Hay & Abby Walker. 2008. Phonological, lexical, and frequency factors in coronal stop deletion in early New Zealand English. Paper presented at Laboratory Phonology 11, Wellington, New Zealand. / Tagliamonte, Sali & Rosalind Temple. 2005. New perspectives on an ol' variable: (t,d) in British English. *Language Variation and Change* 17, 281–302. / Temple, Rosalind. In press. Where and what is (t,d)? In Calamai, Silvia & Chiara Celata, eds. *Advances in Sociophonetics*. Amsterdam: John Benjamins.

Using conversational corpus speech to investigate the impact of speech rate on speech perception performance over the adult life span

Xaver Koch[1] & Esther Janse[1,2,3]

[1]Radboud University Nijmegen - Centre for Language Studies, Nijmegen, the Netherlands
[2]Donders Institute for Brain, Cognition, and Behaviour, Nijmegen, the Netherlands
[3]Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands

x.koch@let.ru.nl, e.janse@let.ru.nl

This study investigates speech perception performance over the adult life span as a function of naturally varying speech rate in spontaneous conversational speech. Research on speech rate effects and aging (e.g., Wingfield et al., 1999) has generally shown that older adults are more impacted by increased speech rates than younger adults (an age by speech rate interaction). Importantly, these studies used lab-recorded and subsequently artificially time-manipulated stimuli, and often time-compressed the test materials to rates beyond the natural range. The present study investigates the same age by speech rate interaction to see whether testing with more ecologically valid materials will lead to the same conclusions as obtained with artificially time-compressed lab-recorded speech. In addition, we include younger, middle-aged, and older adults to be able to investigate adult age in a more continuous way. Further, we investigate which auditory, cognitive and linguistic individual listener characteristics predict the impact of increased speech rate on speech processing.

We set up an eye-tracking study (visual world paradigm) in which participants performed a speeded click task. They had to click on the one out of four words on a visual display they detected in a two-speaker question-answer dyad. This yields two dependent variables: Click-Response-Times (Click-RTs measured from target word offset) and gaze proportions to the target word in an early time window ranging from 200 to 1000 ms after target word onset. The eye tracking data are assumed to reflect the speech processing stage, whereas the Click-RT data represent a composite measure of processing, decision making and motor actions. We tested 102 Dutch participants: 35 younger (Mean age=21), 33 middle-aged (Mean age=50) and 34 older (Mean age=67). Participants were also tested on their hearing, vision, vocabulary knowledge, processing speed and working memory.

Figure 1: Model plot – speech rate effect on Click-RTs

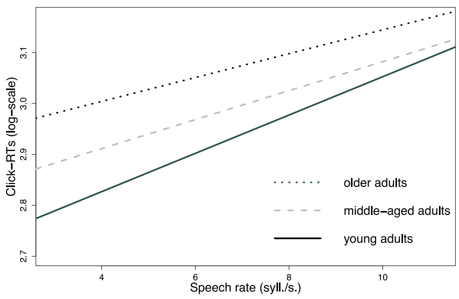
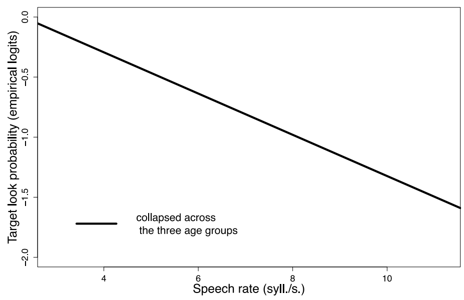


Figure 1: Model plot – speech rate effect on gaze proportions



Linear mixed effect regression analyses of the Click-RTs (Fig. 1) and the gaze proportions (Fig. 2) show that increased speech rate makes sentence processing more difficult in all three age groups. In contrast to earlier findings, however, our Click-RT analyses show that younger adults seem to be more affected by increased speech rates than older and middle-aged adults. Analyses of the gaze proportion data suggest that this discrepancy may be due to task-related differences between the age groups. We assume that there are age group differences in decision and motor response times, rather than in processing itself since no age effect or an age by speech rate interaction effect was found for the proportion of looks to the target word (see Fig. 2). Processing speed, vocabulary knowledge, and hearing differences (particularly among the older adults) are predictors for speech processing performance, but effects are generally stronger for Click RT performance than for gaze performance. The gaze results thus indicate that age differences in speech processing are smaller than suggested by studies using artificially created speech materials (cf. Gordon et al., 2009) and thus stress the need for the use of ecologically valid materials. The data also suggest that individual characteristics predict speech processing differences at speech rates encountered in everyday conversational situations.

References

Gordon, M., Daneman, M., & Schneider, B. (2009). *Comprehension of speeded discourse by younger and older listeners.* Experimental Aging Research, 35, 277-296. / Wingfield, A., Tun, P., Koh, C., & Rosen, M. (1999). *Regaining lost time. Adult aging and the effect of time restoration on recall of time-compressed speech.* Psychology & Aging, 14, 380-389.

Speech reduction in Czech

Alice Kolman[1] & Petr Pollak[2]

[1]Radboud University Nijmegen & Christian University of Applied Sciences CHE

[2] Faculty of Electrical Engineering, Czech Technical University

akolman@che.nl, pollak@fel.cvut.nl

In spontaneous conversations, words often appear with fewer segments than in their citation forms (Ernestus & Warner, 2011). For instance, in American English, the word *yesterday* is often pronounced like *yeshay*. Although this phenomenon of speech reduction is highly frequent, we still know very little about which linguistic factors influence the reduction of segments, and, moreover, the study of reduction has been mainly restricted to Germanic languages (Dutch, English and German) and Romance languages (French and Spanish). The present study contributes to this research by investigating speech reduction in Czech. This West Slavic Language differs from the languages studied so far, among other things, as it contains many consonants (28) and consonant clusters, allows liquids and nasals to be the nuclei of syllables, and is morphologically more complex (it has many derivational and inflectional affixes).

Our study is based on the Nijmegen Corpus of Casual Czech, which was recorded in Prague in November 2008, and consists of 39 hours of casual conversations between 26 groups of three friends (36 male and 42 female speakers, age 19-33). The corpus is orthographically transcribed and the transcriptions are aligned with the speech signal. We studied speech reduction in this corpus by focusing on a number of frequent words and frequent phoneme sequences.

First, we see patterns that have also been observed in other languages. We find clear effects of the word's frequency of occurrence: highly frequent words with non-specific meanings are often highly reduced. For instance, /j/ is often absent in the words /jɛstlɪ/ 'whether' (frequency of 643 in the corpus) and /jɛʃʈɛ/ 'yet, still' (frequency of 1229). The function word /jɛstlɪ/ was often reduced to /ɛstlɪ/, /ɛslɪ/ or /ɛs/, the function word / jɛʃʈɛ / to /ɛʃʈɛ / or /ɛʃ/. In addition, we find that consonant clusters are often reduced. For instance, word-inital /v/ is often absent when followed by a sibilant, and the word-initial consonant cluster /kd/ at the beginning of question words is often reduced to /d/.

Second, Czech also shows clear effects of morphology, which has not been attested for other languages so far. For instance, in the present tense conjugations of the verb /ji:t/ 'to go' (frequency of 144), stem-initial /j/ can be absent when followed by an obstruent. However, if the verb is preceded by the prefixes /nɛ/, /dɔ/, /na/, /prɔ/, /u/, or /za/, /j/ is never absent, independently of the word's frequency of occurrence. This pattern cannot be explained by resyllabification, since /j/ can also be absent when the stem is preceded by the prefixes /pu:/, /pr̝̊ɪ/ or /vɪ/. In these cases, however, the absent /j/ leaves acoustic traces in the duration of the preceding vowel (which is, depending on the prefix, unexpectedly long or short). These fine phonetic cues may help listeners to overcome the absence of a segment, that typically is not absent after a prefix.

The effect of morphology appears to be conditioned by the properties of the following consonant. For instance, stem-initial /j/ can be absent in the root /jmɛnɔvat/ 'to name', in which it is followed by /m/, also when preceded by prefixes like /nɛ/, /pɔ/, /pr̝̊ɛ/ or /vɪ/. In these cases, the absence of /j/ does not affect the duration of the preceding vowels, which strongly suggests that the duration of the preceding vowel does not form a cue to a following absent consonant, if this consonant can be absent under all conditions.

A third interesting topic concerns syllabic consonants, such as /l/ and /r/. We find that the segment's probability to be absent is modulated by the complexity of the resulting consonant cluster. For instance, whereas syllabic /l/ is often absent in /jab**l**kɔ/ 'apple' and /u:p**l**ɲɛ/ 'totally', it is always present in words such as /b**l**bi:/ 'stupid' (where absence of /l/ would result in a sequence of identical consonants). Similarily, syllabic /r/ is always present in words such as /zm**r**zlɪna/ 'icecream', /ʧtv**r**tɛk/ 'Thursday' and /p**r**vni:/ 'first' (even though this latter word has a frequency of 183).

Our study of Czech clearly shows that it is worthwhile to extend the study of reduction to typologically different languages. Speech production models will have to explain how the Czech patterns can emerge, while research in speech comprehension will have to investigate whether listeners use these reduction patterns, and whether they use the cues that are left behind by some consonants in some conditions.

Reference

Ernestus, M., Warner, N. 2011. An introduction to reduced pronunciation variants. *Journal of Phonetics*

On the tail of the Scottish Vowel Length Rule in Glasgow

Tamara Rathcke[1] & Jane Stuart-Smith[2]

[1]University of Kent; [2]University of Glasgow

T.V.Rathcke@kent.ac.uk; Jane.Stuart-Smith@glasgow.ac.uk

This paper draws on a longitudinal corpus of spontaneous Glaswegian speech recorded outside the phonetics laboratory in the 1970s and 2000s. Analyses of vowel duration were carried out to find evidence for a sound change in process involving the famous Scottish Vowel Length Rule (SVLR, Aitken 1981), which has recently been found to be weakening in studies of read speech from several local communities across Scotland (e.g. Llamas et al. 2011).

In contrast to many varieties of English where durational alternations between short and long vowels play a minor role in phonology, traditional accents of Scottish English employ quantity as a distinctive feature in a small part of the lexicon (with minimal pairs like short *crude*, *brood* vs. long *crewed*, *brewed*). SVLR states that vowels bearing primary lexical stress have long allophones before voiced fricatives, /r/ and morpheme boundaries, and short allophones in other contexts (Aitken 1981). In Glaswegian Vernacular, the operation of this rule is restricted to the vowels /i ʉ ai/ (Scobbie et al. 1999).

Interestingly, specifications of SVLR overlap with two well documented timing effects. First, many varieties of English show vowel lengthening in context of a following voiced consonant (e.g. House & Fairbanks 1953). In contrast to the voicing effect, SVLR exhibits more complexity as it includes a detailed specification of the manner of articulation in addition to voicing. The documented tendency for SVLR to become subject to sound change in dialect contact situations with a subsequent simplification towards a more general [±voice] constraint is therefore somewhat unsurprising (Hewlett et al. 1999; Llamas et al. 2011). Second, the morphological conditioning of SVLR also overlaps with the prosodic timing effects of lengthening to demarcate finality within prosodic constituency. Durational ceiling effects are known to occur in quantity languages in situations of combined accentual and utterance-final lengthening (Nakai et al. 2012). The question, then, is whether similar conflicts between SVLR-related lengthening and prosodic lengthening may also act as an additional, system-internal force to weaken the implementation of the SVLR over time.

Conducting the first study to investigate the implementation of SVLR in spontaneous speech, we concentrated on the monophthongs /i ʉ/ only as the durational patterns of /ai/ also involve vowel quality alternations (Scobbie et al. 1999). Our corpus consists of sociolinguistic interviews and informal conversations amongst close friends recorded in non-laboratory settings during the 1970s and 2000s. We compared productions by four speakers groups, two of middle-aged men (1970M and 2000M) and two of adolescent boys (1970Y and 2000Y). According to the information available, each speaker had higher or lower levels of contact with Anglo-English. Extensive manual annotations were carried out on all tokens bearing prominence and containing /i ʉ/ (except preceding /r/). We distinguished between three levels of prominence (stressed, accented and nuclear) and two phrasal positions (medial vs. final). Segmental and morphological environment of each vowel was annotated for SVLR constraints (long vs. short) and consonantal voicing. The dataset was then subjected to a linear mixed effects analysis via a step-wise backward selection of the best-fit model.

Two hypotheses were tested against the data: (1) the *Dialect Contact Hypothesis* (DCH, predicted a consonantal voicing effect in high-contact speakers) and (2) the *Prosodic Timing Hypothesis* (PTH, predicted SVLR-weakening primarily in nuclear and phrase-final positions). The results showed partial support for both predictions with stronger evidence for the impact of prosody. In line with DCH, we found an effect of consonant voicing and level of contact, but only within a three-way interaction with the vowel category ($\chi^2_{(2)} = 9.0$, $p = 0.011$): only /ʉ/, but not /i/, was longer in contexts specified by DCH, with the overall effect being quite tentative (10 ms lengthening, $t = 1.69$, $p = 0.092$). Supporting PTH, a significant three-way interaction of phrasal position with speaker group and SVLR ($\chi^2_{(3)} = 19.55$, $p < 0.001$) indicated that a system-internally driven sound change may be in progress involving a weakening of SVLR in prosodic tails: speakers from the 1970M-group marked phrase-final SVLR-long vowels by 100 ms lengthening while other groups lengthened by 70 ms (1970M/1970Y-comparison: $t = 3.1$, $p = 0.002$) or 50 ms (1970M/2000M-comparison: $t = 2.9$, $p = 0.004$).

We further suggest that the *Prosodic Timing Hypothesis* alone does not entirely explain the main data patterns. Given that the change was observed in speakers who had acquired their vernacular during the time of urban regeneration happening in Glasgow between 1950 and 1970 (i.e. speakers from the 1970Y and 2000M groups), we argue that the results align very well with a key postulate of the social network theory (Milroy and Milroy 1985). Loosening of network structures reduces the stability of vernacular norms, facilitating language change towards a new, system-internally plausible form.

References

Aitken, A.J. 1981. The Scottish Vowel Length Rule. In: Benskin, M., Samuels, M.L. (eds.), *So meny People, Longages and Tonges: Philological Essays in Scots and Mediaeval English presented to Angus McIntosh.* Edinburgh, 131-157. / House, A., Fairbanks, G. 1953. The influence of consonant environment upon the secondary acoustical characteristics of vowels. *J. of Acoust. Soc. of Am. 25(1)*, 105-113. / Llamas, C. et al. 2011. Effects of the Scottish Vowel Length Rule on Vowel quantity in Tyneside English. *Proc. 17th ICPhS Hong Kong*, 1282-85. / Milroy, J., Milroy, L. 1985. Linguistic Change: Social networks and speaker innovation. *J. of Linguistics 21*, 339-84. / Nakai, S. et al. 2012. Quantity constraints on the temporal implementation of phrasal prosody in Northern Finnish. *J. of Phonetics 40*, 796–807. / Scobbie, J., Turk, A., Hewlett, N. 1999. Morphemes, Phonetics and Lexical Items: The Case of the Scottish Vowel Length Rule. Proc. 14th *ICPhS San Francisco*, 1617-20.

Does /t/ produced as [ʔ] involve tongue tip raising? Articulatory evidence for the nature of phonological representations.

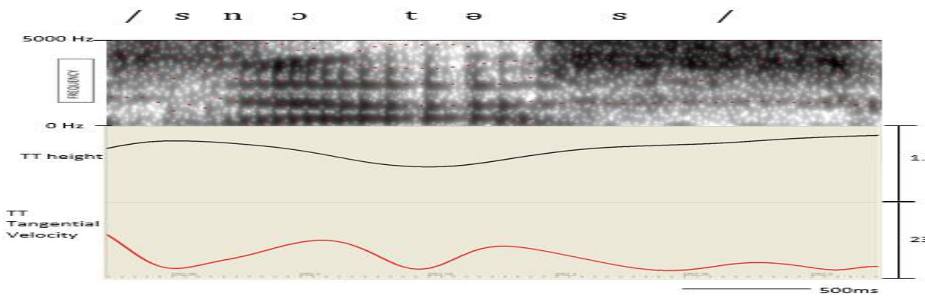
Jennifer Heyward, Alice Turk[a]*, Christian Geng*[b]
The University of Edinburgh[a] & The University of Potsdam[b]
jenny.heyward@gmail.com

This research addresses an ongoing controversy in the field of phonetics and phonology: the nature of phonological representations. Traditional theories propose symbolic, a-temporal representations for contrastive phonological units. In contrast, Articulatory Phonology (Browman & Goldstein 1985 et seq.) proposes spatio–temporal representations in the form of gestures. These approaches make different predictions for the range of possible phonetic variants associated with a phonological unit. We evaluate these predictions by studying the British /t/ phoneme, as produced by speakers of Standard British English and Standard Scottish English, using a corpus of articulatory data from spontaneous and reading speech tasks. By studying tongue tip movement we are able to explore the articulatory components of glottal stop and non-glottal stop /t/ tokens. Articulatory Phonology predicts that tongue tip raising should be present in any variant of /t/ to some degree if time is available (i.e. /t/ is not completely deleted). In contrast, traditional symbolic theories allow for qualitatively different sets of articulators in contextually governed variants of the same phoneme.

We have currently analyzed 1342 intervocalic (non-word-initial) and phrase-final /t/ tokens from the ESPF DoubleTalk corpus (e.g. *pre**tt**y, bu**t** I've, boa**t**…*). Articulate Assistant Advanced software was used to measure the position of a sensor glued to the tongue front (tongue tip sensor) at tangential velocity minima for the preceding and following vowels, and at the tangential velocity minimum associated with peak displacement for /t/ when tongue tip raising was involved. When /t/ was produced without tongue tip raising, the temporal mid-point between the preceding and following vowels was used as the point of tongue tip sensor position measurement. The extent of glottalization was diagnosed visually through spectrographic analysis and the first author performed preliminary auditory and spectrographic analyses to categorize /t/ tokens as [t], acoustically deleted ([Ø]), or as glottal stop ([ʔ]).

The data reveal two types of glottalized /t/ tokens – those in which the glottalization can be attributed to the phrasal domain (either phrase final, or glottalized throughout a relatively long, phrasal domain), and local, short-range glottalization, associated with [ʔ]. /t/ tokens associated with phrasal glottalization frequently exhibit tongue tip raising (147 of the 161 /t/ tokens with phrase-final glottalisation, and 9 of the 31 with longer range glottalisation). Of these, /t/ tokens heard as [t] *always* show tongue tip raising. There are 113 /t/ tokens which, using spectrographic and auditory analyses, are categorized as clear glottal stops. Of these, only two show any evidence of tongue tip raising. The remainder 111 show no signs of tongue tip movement for /t/; Figure 1 shows an example:

Figure 2: The spectrogram and EMA trace for part of 'it's not a squared…' produced with a glottal stop by speaker R0033_cs6.



These results are consistent with symbolic abstract phonological representations: different allophones are realized with qualitatively different sets of phonetic correlates – [t], with tongue tip raising, and [ʔ] without tongue tip raising and a localized glottal constriction. These results may be difficult to account for in theories where /t/ is always associated with tongue tip raising. Currently, in Articulatory Phonology, for example, tongue tip raising would be expected as long as there is gestural activation for /t/; any gestural blending with adjacent vowels would compromise tongue body movement so that the tongue tip would move further in compensation to achieve the /t/ target.

References:

Browman, C.P., & Goldstein, L. (1985). Dynamic modeling of phonetic structure. In V. A. Fromkin (Ed.), *Phonetic Linguistics* (pp. 35-53). New York: Academic Press. / Scobbie, J.M., Turk, A., Geng, C. , King, S., Lickley, R.J., & Richmond, K. The Edinburgh Speech Production Facility Doubletalk corpus. *INTERSPEECH 2013*, 764-766.

A corpus based investigation of the contrast between French rise-fall and rise via wavelet based functional mixed models

Cristel Portes[1], Leonardo Lancia[2]

1. Aix-Marseille Université, CNRS, LPL, UMR 7309, 13604, Aix-en-Provence, France.
cristel.portes@lpl-aix.fr, leonardo_lancia@eva.mpg.de

Proposed inventories of French intonational contours are rarely based on a systematic and quantitative analysis of corpus data (Delattre 1966, Di Cristo & Hirst 1996, Mertens 2008, Rossi 1999). On the other hand, with the exception of Post's work (Post 2000), experimental studies have focused on the structure of the lower level prosodic constituent (the accentual phrase), more than on the inventory of contours at the higher level of the intonational phrase (Astesano 2007, Jun & Fougeron 2000; Welby 2006). A possible explanation is that some common contours are difficult to elicit in laboratory experiments, requiring specific conversation contexts and situations. However, a recent trend in phonology aims at exploring more natural corpus data thanks to the availability of new analytical tools (Coleman et al., 2011; Cole & Hasegawa-Johnson, 2012). This approach permits the observation of contours hard to elicit in the laboratory. Moreover it may give access to a better understanding of the full range of variation in the use of contours by speakers. French rise-fall (RF) is one of these contours. Delattre (1966) called it "intonation d'implication" (implication tune) and described it as a rise through the penultimate syllable of the utterance with the f0 peak occurring on the final (primary accented) syllable and then falling slightly to the end. More recently, Post (2000) proposed an autosegmental-metrical account of French contours inventory where this contour was coded LH*L%. Its contrast with the simple rise LH*H% (hereafter R) was proposed to rely on an alignment difference between the rising parts which is delayed on the final syllable for R, and on the addition of a contrastive falling part for RF. The height of both f0 maxima was supposed to be similar. Conversely, Rossi (1999) conceived RF as derived from the R by an increased pitch height. Moreover, while Mertens (2008) argued for a phonological contrast between HB and HB- based on differences in the final height of RF, Ladd (2008: 122) suggests that the fall from the peak "seldom, if ever, reaches the bottom of the range". The main aim of the present study was to compare the phonetic realization of RF and R in a corpus of naturally occurring data in order to decide whether the alignment difference or the height difference is more relevant to the contrast. A 45 minutes radio debate between six interlocutors was recorded, orthographically transcribed and automatically labeled for phones. Two experts of French intonation were asked to auditorily identify all utterance final contours in the corpus using Delattre (1966)'s inventory. A total of 114 "intonations d'implication", i.e. RF and of 460 "continuations majeures", i.e. R was identified; 62 *consensual* items of RF were identified by both experts and 52 *ambiguous* items by one of the two (the other one identifying R). We selected 117 consensual items of R for the comparison. We compared the f0 curves of consensual RF, ambiguous contours (Amb) and consensual R on the nucleus of the final syllable. To this aim we used a wavelet based functional mixed model (Morris & Carroll, 2006) in which the dependent variable was the f0 curve. The fixed factors were the nature of the contour (RF, Amb and R coded trough a successive difference contrast), the presence vs absence of a coda and the interaction between these two factors. Two random intercepts were included: speaker identity and vowel identity; plus a speaker specific random slope for each of the predictors. We also measured f0 mean height on the nucleus of the penultimate syllable which was expected to be higher for RF than for R due to the alignment difference. This hypothesis was tested trough a mixed model with the same predictors and random effects structure adopted for the other model, with the exclusion of the random intercept for vowel identity. Results show that: 1) RF, Amb and R contours have different alignment of the f0 peak (beginning, mid and end of the final nucleus respectively); 2) in the coda condition, the f0 scaling at the end of the final syllable is higher for R than for Amb which is higher than RF), and RF has a higher f0 peak than the other two, 3) mean f0 on the nucleus of the penultimate is higher for RF than for R. These results confirm the alignment difference proposed by Post (2000) and show that only the coda condition support Rossi (1999)'s proposal of a higher RF. They also show the usefulness of functional mixed models to asses the effects of several factors on f0 trajectories without a priori assumptions on their shape.

References

Astésano, C., Bard, E., & Turk, A. 2007. Structural influences on Initial Accent placement in French. *Language and Speech* 50, 423-446. / Cole, J., Hasegawa-Johnson, M. 2012. Corpus phonology with speech resources. In: Handbook of Laboratory Phonology. Fougeron C., Huffman M. (Eds.), Oxford University Press, 431-440. / Coleman, J., Liberman, M., Kochanski, G., Burnard, L., Yuan, J. 2011. Mining a Year of Speech. In: Proc. Of VLPS 2011, University of Pennsylvania. / Delattre, P. 1966. Les dix intonations de base du français. *French Review* 40, 1-14. / Di Cristo, A., Hirst, D. 1996. Vers une typologie des unités intonatives du français, In: Société française d'acoustique (ed.), *16émes Journées d'Etude sur la parole*. Avignon, 219-222. / Jun, S.-A., Fougeron, C. 2000. A phonological model of French intonation, in Botinis, A. (ed.) *Intonation : Analysis, Modeling and Technology*, Dordrecht : Kluwer Academic Publishers, 209-242. / Mertens, P. 2008. Syntaxe, prosodie et structure informationnelle : une approche prédictive pour l'analyse de l'intonation dans le discours. *Travaux de Linguistique* 56(1), 87-124. / Morris, J. S., & Carroll, R. J. (2006). Wavelet-based functional mixed models. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 68(2), 179-199. / Rossi, M. 1999. *L'intonation, le système du français : description et modélisation*. Paris : Ophrys. / Post, B. 2000. *Tonal and phrasal structures in French intonation*. The Hague: Holland Academic Graphics. / Welby, P. (2006). French intonational structure: Evidence from tonal alignment. *Journal of Phonetics,* 34(3), 343–371.

A New Corpus of Colloquial Korean and its Applications

Kevin Tang, Brent de Chene
University College London, Waseda University
kevin.tang.10@ucl.ac.uk, dechene@waseda.jp

Motivation: Speech produced outside the phonetics laboratory provides ecological validation for experimental findings. Using a newly constructed subtitle corpus, we present here a corpus-based approach to modeling variations in spontaneous speech. While everyday spontaneous speech is inexpensive to collect from field work and digital media, the transcription time and cost remain obstacles to creating a large corpus, capable of capturing a wide range of sensitive linguistic variations. The existing corpora of Korean are too small to be representative of the language. For instance, the spoken-transcripts in the 21st Century Sejong Corpora (Kim and Kang, 1998) provide 5.2 million words and the ETRI database (2006) consists of 24,300 sentences read by a single speaker. The lack of a suitable corpus motivated us to construct our own corpus of conversational Korean.

SUBTLEX, a method of using film subtitles to construct corpora, was developed by New et al. (2007) with French, and has since been applied to many other languages (see, e.g., Brysbaert and New, 2009; Keuleers, Brysbaert, and New, 2010). Crucially these film subtitle frequencies have been proven to be excellent predictors of behavioural task measures such as reaction times in lexical decision tasks. This is primarily due to their spoken register, as they are essentially transcribed spoken speech, as well as their large corpus size. Following this method, we set out to compile a subtitle corpus of Korean, SUBTLEX-KR.

Method: A Korean subtitle website was datamined, and the files preprocessed to remove irrelevant information. Non-Korean files were filtered by applying a language detection model (Shuyo, 2010). Subtitle files were screened for duplicates using the Kullback–Leibler divergence (Kullback and Leibler, 1951). The resulting corpus contained 90 million word tokens and 3.6 million word types, where "word" = *eojol*, a unit between spaces in Korean writing.

Applications: SUBTLEX-KR is useful in investigating a wide range of Korean phonological phenomena that involve variation, both when that variation represents regularization in progress and when it represents alternative resolutions of conflicting forces. Here we will briefly treat three such phenomena. The first is variation in the realization before vowel-initial clitics of noun stems that end historically in marked obstruents (/p^h, t^h, c^h, c, k^h, k'/) or clusters (/ps, ks, lk/, etc.). For many such stems, the historically expected prevocalic allomorph varies with an allomorph that coincides with the prepausal and preconsonantal form (for stems ending in coronals, with a form resulting from a productive rule that takes *t* to *s* prevocalically at the end of a noun stem). After noting evidence that, as widely assumed, underlying forms of noun stems have been reanalyzed as coinciding with prepausal or isolation forms, we will introduce two possible interpretations of this variation. On one of them, variation is due to stochastic rules that mirror lexical statistics (Jun 2010); on the other, variation is the result of ongoing elimination of irregular allomorphs, with concomitant expansion of the range of the corresponding default forms. These two interpretations make different predictions about which stems should vary and which should be stable. While existing corpora fail to supply evidence relevant to the choice between them, we will show that SUBTLEX-KR provides clear evidence in favor of the interpretation on which variation results from ongoing regularization. The second phenomenon we will investigate is variation involving the residual vowel harmony alternation *a*∼*ə* in verbal inflectional suffixes. We will compare results from SUBTLEX-KR with earlier results from other corpora (Hong 2008, Kang 2012) and present an interpretation of this case as well in terms of ongoing regularization. Finally, we will consider variation in the epenthesis of /ɨ/ after stop-final English loanwords and show that SUBTLEX-KR allows us to test the conclusions of Kang (2003) with regard to this phenomenon.

References

Brysbaert, M., New, B. 2009. Moving beyond Kučera and Francis: a critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. In: *Behavior Research Methods* 41.4, 977–990. / Electronics and Telecommunications Research Institute (ETRI). 2006. Database of conversational sentences for speech synthesis. http://slrdb.etri.re.kr/DBSearch/Voice.asp. / Hong, S-H. 2008. Variation and exceptions in the vowel harmony of Korean suffixes. *Journal of Studies in Language* 24:405-428. / Jun, J. 2010. Stem-final obstruent variation in Korean. *Journal of East Asian Linguistics* 19:137-179. / Kang, H. 2012. *Diachrony in synchrony: Korean vowel harmony in Verbal Conjugation.* Dissertation, Stony Brook University. / Kang, Y-J. 2003. Perceptual similarity in loanword adaptation: English postvocalic word-final stops in Korean. *Phonology* 20:219–273. / Keuleers, E., Brysbaert, M., New, B. 2010. SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. In: *Behavior Research Methods* 42.3, 643–650. / Kim, H-G., Kang, B-M. 1998. 21st Century Sejong Project-Compiling Korean Corpora. In: *Development* 1999, 2000. / Kullback, S., Leibler, R.A. 1951. On information and sufficiency. In: *The Annals of Mathematical Statistics* 22.1, 79–86. / New, B. et al. 2007. The use of film subtitles to estimate word frequencies. In: *Applied Psycholinguistics* 28.4, 661–677. / Shuyo, N. 2010. Language Detection Library for Java. url: http://code.google.com/p/language-detection/.

Phonological CorpusTools: A Free, Open-Source Tool for Phonological Analysis

Scott Mackie, Kathleen Currie Hall, Blake Allen, Michael Fry, Michael McAuliffe
University of British Columbia
Corresponding authors: mackie@email.com, kathleen.hall@ubc.ca

In this poster, we introduce and demonstrate *Phonological CorpusTools* (PCT), a new freely available open-source tool for doing phonological analysis on transcribed corpora. PCT is intended to be an analysis aid for researchers who are specifically interested in investigating the relationships that may hold between individual sounds in a language. There is an ever-increasing interest in exploring the roles of frequency and usage in understanding phonological phenomena (e.g., Bybee 2001, Ernestus 2011, Frisch 2011), but many corpora and existing corpus-analysis software tools are focused on dialogue- and sentence-level analysis, and/or the computational skills needed to efficiently handle large corpora can be daunting to learn.

PCT is designed with the phonologist in mind and has an easy-to-use graphical user interface that requires no programming knowledge. It specifically includes the following capabilities:

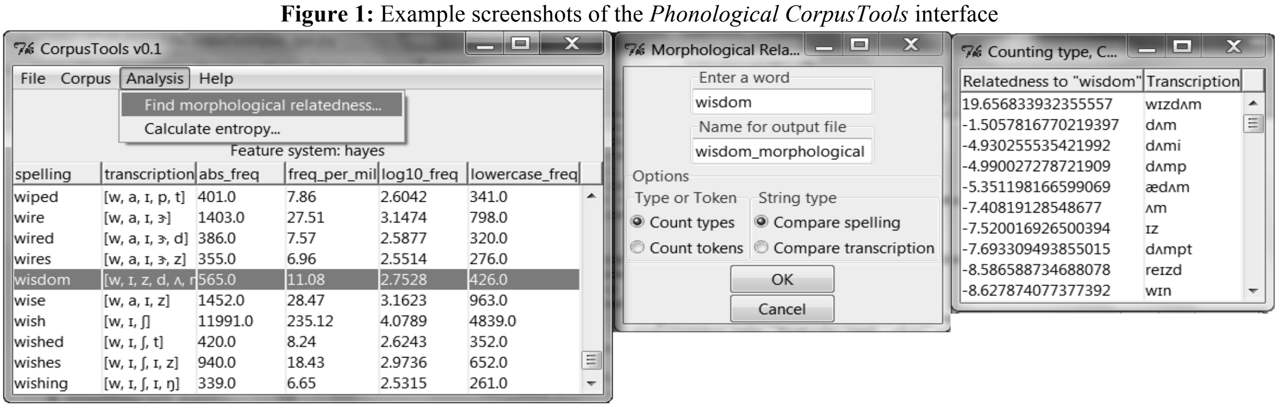
- Calculation of functional load of individual pairs of sounds, defined at either the segment or feature level (cf. Surendran & Niyogi 2003; Wedel, Kaplan, & Jackson 2012);
- Calculation of the extent to which any pair of sounds is predictably distributed given a set of environments that they can occur in, as a measure of phonological contrastiveness (cf. Hall 2009, Hall & Hall 2013);
- Calculation of the extent to which pairs of words are similar to each other using either orthographic or phonetic transcription, which can be used as a proxy for morphological relatedness (cf. Frisch et al. 2004, Khorsi 2012);
- Calculation of the frequency with which two sounds alternate with each other, given a measure of morphological relatedness (cf. Silverman 2006, Lu 2012);
- Calculation of the phonetic similarity between sounds or words, based on alignment of logarithmically spaced amplitude envelopes derived from sound files (cf. Lewandowski 2012).

The software can make use of pre-existing freely available corpora (e.g., the IPHOD corpus (Vaden et al. 2009)), or a user may upload his or her own corpus in either of two formats. First, lexical lists with transcription and frequency information can be directly uploaded. Second, raw running text (orthographically and/or phonetically transcribed) can be uploaded and turned into lexical lists in columnar format for subsequent analysis. Raw sound files accompanied by Praat TextGrids (Boersma & Weenink 2014) may also be uploaded for analyses of phonetic similarity.

Analysis can be done using built-in feature charts based on Chomsky & Halle (1968) or Hayes (2009), or a user may create his or her own specifications by either modifying these charts or uploading a new chart. Feature specifications can be used to pull out separate "tiers" of segments for analysis (e.g., consonants vs. vowels, all nasal elements, tonal contours, etc.). Analysis can be done using type or token frequency. All analyses are presented on screen and saved to plain .txt files.

The software is written in Python 3.3, and users are welcome to add on other functionality as needed. The software works on any platform that supports Python (Windows, Mac, Linux). As part of the poster presentation, we will provide information on accessing the software and demonstrations of its utility.

Figure 1: Example screenshots of the *Phonological CorpusTools* interface



References

Available from: http://anonymousreferences.wordpress.com/2014/01/31/abstract-phonological-corpustools/

Schwa reduction in spontaneous infant-directed speech

Mybeth Lahey [1,2,3] *& Mirjam Ernestus* [2,1]

[1] Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands; [2] Radboud University Nijmegen, Nijmegen, The Netherlands; [3] International Max Planck Research School for Language Sciences, Nijmegen, The Netherlands

mybeth.lahey@mpi.nl, m.ernestus@let.ru.nl

Spontaneous speech between adults is abundant with pronunciation variation related to connected speech processes. For example, in conversational speech, the schwa in the English word *suppose* may be shortened or even deleted, as in *s'pose*. The question is how infants acquire these reduced variants. Infants acquire the phonology of their language simply by listening to the speech around them. Their primary language input consists of infant-directed speech (van de Weijer, 1998). In the present corpus study, we investigated whether such reduced forms are present in the input to preverbal infants. We focused on a common reduction process in Dutch: schwa reduction in unstressed syllables.

Infant-directed speech is generally characterized by a slower speaking rate, larger vowel triangles, a higher pitch and exaggerated pitch contours. However, a few studies have shown that infant- and child-directed speech is replete with connected speech processes such as reduction. Importantly, these studies typically focused on read speech instead of spontaneous speech (Dilley et al., 2014), on speech directed to toddlers (Shockey and Bond, 1980) or on a limited set of speakers or word types (Buchan and Jones, 2013; Lahey and Ernestus, 2014). The question thus remains open whether infants are confronted with reduced pronunciation variants on a large scale in the naturalistic input at the earliest stages of language acquisition. In the present study, we investigated acoustic reduction in content words in infant-directed speech.

We investigated schwa reduction in initial unstressed syllables of Dutch content words (e.g. *gezien* 'seen' and *vergiet* 'colander'), directed to 11- and 12-month-old infants. We analyzed the presence and duration of schwa in word tokens extracted from an extensive corpus (about 39 hours of speech) of 56 Dutch caregivers interacting with their children. Our sample contained over 100 word types. We studied the effects of speech rate, utterance length, word predictability and individual differences between caregivers, such as speaker's sex and age, on reduction.

We found that schwa was absent in 21% of word tokens. When schwa was present, its duration ranged from 10 to 136 ms. These findings show that infants are confronted with considerable variation in the phonemic content of words and in schwa duration. Furthermore, we found several factors affecting absence and shortening of schwa. Speakers under 36 years of age uttered more words with absent schwa than older speakers. For schwa duration, we also found that younger speakers tend to produce shorter schwas. Surprisingly, our results show that female caregivers produced shorter schwas (mean 43 ms) than male caregivers (mean 55 ms). Also, schwas were shorter in utterances containing more words and when local speech rate (in syllables per second) was higher. We found no effect of frequency of occurrence of the target word.

This study is one of the first to show that acoustic reduction is present in a large number of word types in naturalistic speech to preverbal infants. Furthermore, we have shown that differences between speakers, utterance length and speech rate play a role in schwa reduction in speech to preverbal infants. This is consistent with findings on schwa reduction in adult-directed speech (e.g. Hanique et al., 2013). We conclude that, around the time of their first birthday, infants are confronted with segmental reductions in content words on a large scale. These reductions are highly similar to those in conversational speech between adults. Our results thus show that infants can acquire knowledge about reduction in spontaneous speech from early in linguistic development.

References

Buchan, H., Jones, C. 2013. Phonological reduction in maternal speech in northern Australian English: change over time. *J. Child Lang.*, 1- 31. / Dilley, L.C., Millett, A.L., McAuley, J.D., Bergeson, T.R. 2014. Phonetic variation in consonants in infant-directed and adult-directed speech: the case of regressive place assimilation in word-final alveolar stops. *J. Child Lang.*, 41(1), 155-175. / Hanique, I., Ernestus, M., Schuppler, B. 2013. Informal speech processes can be categorical in nature, even if they affect many different words. *J. Acoust. Soc. Am.*, 133(3), 1644-1655. / Lahey, M., Ernestus, M. 2014. Pronunciation variation in infant-directed speech: phonetic reduction of two highly frequent words. *Lang Learn Dev.* in press. / Shockey, L., Bond, Z.S. 1980. Phonological processes in speech addressed to children. *Phonetica*, 37, 267–274. / van de Weijer, J. 1998. *Language input for word discovery.* Nijmegen: Max Planck Institute for Psycholinguistics.

Uptalk in semi-spontaneous and scripted speech

Amanda Ritchart[1] *and Amalia Arvaniti*[2]
[1]University of California, San Diego [2]University of Kent
aritchart@ucsd.edu a.arvaniti@kent.ac.uk

Rising melodies used with statements, commonly referred to as *uptalk*, are attested in English as spoken in Australia, New Zealand, North America and the UK (e.g. Fletcher et al., 2006, and references therein). One of the varieties often associated with uptalk is Californian English, though uptalk in this variety has not been studied in detail. Here we present data from Southern California uptalk as manifested in semi-spontaneous and scripted speech. The former data revealed systematic regularities in the form and function of uptalk, while scripted speech allowed for the testing of specific hypotheses about the representation of uptalk leading to the development of the present analysis.

The study included 11 male and 12 female Southern Californian speakers of all socioeconomic classes (SEC). Thirteen speakers were monolingual, while the others were bilingual in English and one of the following: Vietnamese, Japanese, Armenian, Assyrian, Spanish, Cantonese. Data were elicited using four tasks: (1) map task with local landmarks and the participant as leader; (2a) reading of the transcript of a sitcom scene; (2b) unscripted retelling of the sitcom scene (henceforth *clip retell*); (3) specially designed questions and statements. The analysis presented here is based on the data from (1), (2b) and (3), as (2a) did not produce many instances of uptalk. For the map task and clip retell each instance of uptalk was classified into one of the following discourse functions: (i) statement, (ii) question, (iii) confirmation request (e.g. indirectly asking the interlocutor if they are following) and (iv) floor holding (which included continuations and listing). Two main measurements were taken: the excursion of the pitch rise (in Hz), and the distance (in ms) of the pitch rise start from the onset of the last stressed vowel in the utterance. The data were analyzed using linear mixed-effects models with Speaker as random intercept.

The map task and clip retell revealed differences in the form and frequency of uptalk based on discourse function, task and gender but showed no effect of SEC or linguistic background. Specifically, uptalk was used more frequently by women than men with the differences being due to floor holding: women used uptalk twice as much for this function (59% vs. 28%), while use of uptalk with statements did not differ by gender (17% for women vs. 16% for men). Uptalk was used in all instances of questions and confirmation requests in the map task and clip retell, and was more frequent in the former than the latter (34% vs. 20%). However, task and discourse function did not interact suggesting that the greater frequency of uptalk in the map task was due to the interactive and cooperative nature of the task rather than the greater use of questions or confirmation requests. Differences were also found with respect to form: uptalk used with questions showed an earlier pitch rise start and greater pitch excursion than uptalk used with statements. In both statements and questions, women showed later alignment and greater pitch excursion than men, suggesting that overall women preferred short sharp rises, while men preferred long and shallow ones. Confirmation requests presented both the pattern found with statements and that found with questions, reflecting their dual role in discourse. Floor holding, on the other hand, was often realized not with uptalk proper (actual pitch rises) but, rather, with the use of high plateaux, which impressionistically resemble rises (cf. Clopper & Smiljanic, 2011). Taken together these results suggest that the melody typically used for rising statements is L* L-H% and that used with questions is L* H-H%, while floor holding is realized as L* H-L%.

Alignment averages, however, hide gender differences which have implications for the analysis presented above: specifically, while in statements both men and women started the pitch rise after (or late in) the nuclear vowel, women started the rise within this vowel in questions as well, while men, on average, started just before it. This difference could indicate a preference for a L* accent for the women and a L+H* accent for men rather than a uniform L* choice. Clarification comes from the scripted questions in which the position of the nuclear syllable was controlled so as to appear early or late in the utterance; cf. (1) *Did you eat the WATERMELON?* vs. (2) *Did you eat the LIME?* For both men and women, the nuclear syllable in the scripted questions exhibited a marked dip and no rise when sufficiently early, as in example (1). This pattern confirms that the accent in questions is indeed a L* and that the rise on the nuclear syllable, when present, is due to tonal crowding and concomitant coarticulation, which appears to be more extensive in the data of the male speakers. The patterns described above document the use of uptalk in Southern Californian English and show that this variety makes a principled distinction between uptalk for statements and questions. The distinction is typically realized as a choice of tune. Finally, these results illustrate the complementarity of spontaneous and controlled data: the former allow us to understand the extent of possible variability in realization but the latter are also needed to confirm or refute hypotheses developed on the basis of the more variable (semi-)spontaneous speech.

References

Clopper, C. G., & Smiljanic, R. 2011. Effects of gender and regional dialect on prosodic patterns in American English. *Journal of Phonetics* 39, 237-245. / Fletcher, J., Grabe, E. & Warren, P. 2006. Intonational variation in four dialects of English: the High Rising Tune. In: Jun, S. (ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press, 390-409.

Variability in the phonology of child-directed speech: evidence from a new naturalistic corpus

Sam J. Green
University College London
sam.green.10@ucl.ac.uk

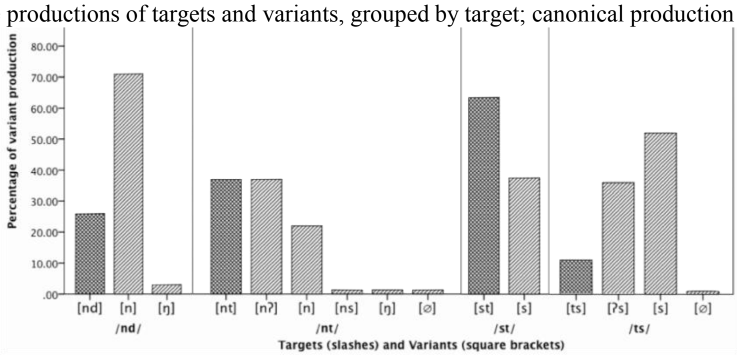
Investigations into first language acquisition of phonology that take language input (and the various statistical distributions of sounds) into consideration as a model that the child learns from, often use citation/dictionary phonetic forms. This approach neglects some – albeit mixed – findings that the speech directed to infants and children (CDS) may be less 'hyperarticulated' than might be assumed (Cristia and Seidl, 2013; Dilley et al., 2013). Previous studies have taken place in experimental settings, using tasks to elicit CDS from parents; often either reading tasks or object naming to elicit certain sounds and sound combinations (Foulkes et al., 2005; Stoel-Gammon, 1984).

The current study is different in that it uses naturalistic data to describe the nature of CDS by utilising a new longitudinal corpus of spontaneous recordings of dyadic interaction between three monolingual British English speaking mothers and their children (children's ages: 0;11 to 2;11), transcribed using Phon (Rose et al., 2006). Whilst some previous studies have focused on vowel production in CDS (e.g. Kuhl et al., 1997), the current study focuses on consonant complexity at the end of words: word-final bi-consonantal clusters. This is being carried out as a first step in a project assessing the acquisition process of these clusters. These structures are understudied in acquisition (Demuth, 2011), hence the motivation to initially investigate the ambient source of children's information regarding them. The study takes the form of a frequency analysis of cluster 'alternations', or variants (consisting of deletion and substitution processes, and combinations resulting from assimilation and coarticulation), compared with expected targetlike – or canonical – productions by adults. Subsequently, the phonological environments in which variants occur are described, as well as patterns where consistent environmental factors are not apparent, and variability over time.

Figure 1 displays a sample of the variability produced by one mother in four high frequency clusters within a single one-hour session, where some variants from the canonical occur as often as 70% of the time. Certain types of processes occur more frequently than others in CDS. In addition, the amount of variability in more frequent clusters is found to be high. The result of this is an altered account of the frequency of canonically produced segmental input to children, at least when considering consonant segments in strings. Explanations of variability potentially relating to cluster types, situational context, and phonological and morphological structure are considered.

The implication, then, is that children's phonology input is not as 'targetlike' as might be sometimes assumed. If children learn considerably from the input, and the input is not perhaps consistent or in the same form as citation forms, then children might not be expected to 'accurately' produce those sounds. This subsequently implies that measures of accuracy and development in previous studies may not be measuring as accurately as they could if specific measurements of the language input were to be taken into account.

Figure 1: Four word-final bi-consonantal cluster Targets and their Variant production (n=734) by one mother in a one hour session (bars show percentage productions of targets and variants, grouped by target; canonical production in cross-hatched bars)



References
Cristia, A. and Seidl, A. (2013). The hyperarticulation hypothesis of infant–directed speech. *Journal of Child Language*. / Demuth, K. (2011). The acquisition of phonology, in J. Goldsmith, J. Riggle and A. Yu, (Eds.) *The Handbook of Phonological Theory*. Blackwell, Malden, MA, pp. 571–595. / Dilley, L. C., Millet, A. L., McAuley, J. D. and Bergeson, T. R. (2013). Phonetic variation in consonants in infant-directed and adult-directed speech: the case of regressive place assimilation in word-final alveolar stops. *Journal of Child Language*. / Foulkes, P., Docherty, G. and Watt, D. (2005). Phonological variation in child-directed speech. *Language 81*(1), 177–206. . Kuhl, P., Andruski, J., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., Stolyarova, E. I., Sundberg, U. and Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science 277*, 684–686. / Rose, Y., MacWhinney, B., Byrne, R., Hedlund, G., Maddocks, K., O'Brien, P. and Wareham, T. (2006). *Introducing phon: A software solution for the study of phonological acquisition*, in D. Bamman, T. Magnitskaia and C. Zaller (Eds.), 'Proceedings of the 30th Boston University Conference on Language Development', pp. 489–500. / Stoel-Gammon, C. (1984). Phonological variability in mother-child speech. *Phonetica* 41, 208–214.

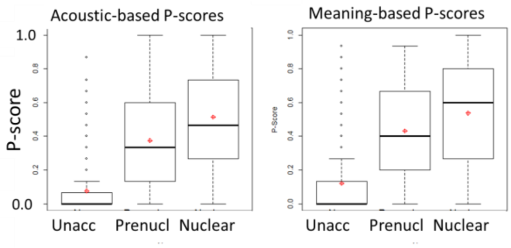
The perception of phrasal prominence in conversational speech

Jennifer Cole, José I. Hualde, Tim Mahrt, Christopher Eager, Suyeon Im
University of Illinois
{jscole|jihualde|tmahrt2|eager2|sim16}@illinois.edu

Prominence relations among words in a phonological phrase determine accent placement and affect the acoustic realization of words. Since Bolinger's (1958) discovery of the role of pitch in cueing accentual prominence in English, a tension has arisen between two strategies: equating accent with pitch excursions (pitch accent) and relying on the perception of trained transcribers for identifying accented words. Both strategies are notoriously fraught with problems. Uncertainty in the identification of prominence arises due to variability in the acoustic cues to prominence, including F0, but may also reflect variability in the phonological context of the utterance (e.g., rhythmic factors), in the status of prominence as marking focus, or in factors affecting lexical accessibility (Calhoun, 2010; Cole, et al., 2010).

Here we introduce a methodology for the study of prominence perception in spontaneous speech, with results for American English. We compare responses of non-expert listeners judging the prominence of words within discourse fragments under two instructions: mark prominence based on acoustic criteria, or based on criteria related to discourse meaning. Prosodic transcription was done using a web interface, allowing the inclusion of participants from the broader community, and was performed in real-time based only on the listener's auditory impression of prominence. 16 native English speakers listened to excerpts of recorded conversations from the Buckeye corpus (16 speakers, 925 words total) and marked a transcript appearing on the screen by selecting words judged as acoustically prominent, or as important for conveying the speaker's intended meaning. Prominence labels from all transcribers were aggregated to produce a P-score for each word, representing the proportion of transcribers who judged that word as prominent. P-scores were separately computed, for each word, under acoustic- and meaning-based transcriptions. The same discourse fragments were also prosodically labeled by two of the authors using ToBI conventions, and ToBI labels were compared with P-scores to establish the relationship between prominence as judged by expert and non-expert listeners.

Linear and general regression models tested the contribution of acoustic cues (word duration, normalized log F0 and intensity), part-of-speech, and word frequency to prominence judgments based on P-scores (acoustic- and meaning-based) and ToBI labels. The main findings are summarized here. (1) *Naïve prominence judgments are a good approximation of expert prosodic transcription.* P-scores are significantly different for words ToBI-labeled as Accented vs. Unaccented, for both acoustic-based P-scores (Wilcoxon W=14329.5, p<.0001) and meaning-based P-scores (W=11812, p<0.0001), and both types of P-scores are strongly predictive of the Accented/Unaccented distinction (likelihood-ratio-test, Pr(>χ2) <0.0001). But only P-scores based on acoustic criteria differentiate ToBI pitch accent labels: Higher P-scores with L+H* and H+!H*, and lower P-scores with H*, !H*, and L*. (2) *Listeners weigh acoustic and meaning-based factors differently in judging prominence.* All acoustic and meaning-based factors are significantly correlated with P-scores under both transcription criteria, but acoustic factors are more strongly predictive of prominence under acoustic criteria, while lexico-syntactic factors are more strongly predictive under meaning-based criteria. (3) *Prominence is higher when judged by meaning criteria compared to acoustic criteria*, but primarily for words that have a prenuclear ToBI accent. (4) *Nuclear accented words are more likely to be judged as prominent* (with higher P-scores) under both transcription criteria, indicating greater perceptual salience. See figure below. Nuclear accented words also exhibit enhanced acoustic cues to prominence.



These findings have important consequences for our understanding of the phonological notion of phrasal prominence. Our research shows that the perception of prominence in discourse is somewhat malleable, as listeners may differentially attend to acoustic cues or to meaning. There is a remarkable convergence between ToBI labeling of accents by trained transcribers and prominence as judged by non-expert listeners, showing that non-expert prominence judgments, when aggregated, provide a solid, reliable and scientifically-sound basis for identifying prominence.

References. Bolinger, D. 1958. A theory of pitch accent in English. Word, 14, 109-49. / Calhoun, S. 2010. The centrality of metrical structure in signaling information structure: A probabilistic perspective. *Language, 86,* 1-42. / Cole, J., et al. 2010. Signal-based and expectation-based factors in the perception of prosodic prominence. *Laboratory Phonology, 1,* 425–452.

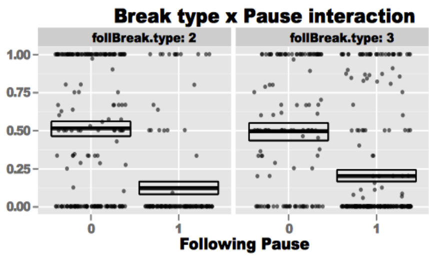
The influence of prosodic boundaries on high vowel devoicing in Japanese

Oriana Kilbourn-Ceron
McGill University
oriana.kilbourn-ceron@mail.mcgill.ca

There is a process in Japanese which devoices high vowels when they are found between two voiceless consonants or a voiceless consonant and pause, termed "high vowel devoicing" (HVD). The true nature of HVD remains controversial. Where should HVD be situated along the phonology-phonetics spectrum? In the Tokyo dialect, HVD occurs even in slow speech (Varden, 1998, 2010), and sounds odd if omitted, consistent with a phonological assimilation analysis (e.g. Tsuchida, 1997). However, not every vowel in an HVD environment is devoiced. The aim of this paper is to model to phonological factors which may shape the pattern of HVD underapplication, and determine their effect relative to phonetic factors. While there is an extensive literature on the factors which inhibit HVD, this study is novel in two ways: it is based on data from spontaneous speech, drawn from the Corpus of Spontaneous Japanese (Maekawa et al., 2000), and second, it focuses on "single devoicing environments" (SDE), which, as noted by Maekawa and Kikuchi (2005), show surprisingly high rates of underapplication in spontaneous speech.

When vowels in adjacent syllables are eligible to devoice, a "consecutive devoicing environment" (CDE), only one of the two actually devoices in most cases. The choice of which vowel to devoice is shows variability, even among identical phone sequences. The factors which influence this choice have been the primary focus of most previous studies on HVD, which have largely employed controlled production experiments (Tsuchida, 1997, 2001; Kondo, 1997; Varden, 1998, 2010). It is implied that outside CDEs, in SDEs, HVD application is almost exceptionless, yet this is not borne out in spontaneous speech. In this study, we focus on one specific factor which previous studies have found to disfavor HVD: morpheme boundaries. By restricting our data to tokens found at the right edge of a morpheme, we can compare two variables: the presence of a following pause, and the strength of the following prosodic boundary (as annotated in the corpus using the J_ToBI schema). If it is revealed that prosodic break type has a significant effect independent of the phonetic environment created by a following pause, it would suggest that HVD interacts with phonological objects such as prosodic boundaries, and so must be represented at the phonological level. This conclusion would be strengthened by including in the model factors identified in previous studies for CDE devoicing.

A corpus of 7766 tokens produced by 201 different speakers was used to fit a mixed-effects logistic regression with



by-speaker and by-morpheme random effects, with maximal random effect structure (Barr et al. 2013). Surprisingly, the model fitted predicts that the base probability of devoicing is very low, only 8.40%. The figure above illustrates the observed devoicing percentage for the factors of greatest interest in this study. A physical pause following a token significantly decreases the odds of devoicing by a factor of 0.047 (b = -3.049, p < 0.001), while a stronger prosodic break type increases the odds by a factor of 1.69, but this is only marginally significant (b = 0.524, p = 0.37). The interaction term between these two predictors reveals that the pause has a much stronger inhibitory effect at a weaker break (b = 1.162, p = 0.005). This suggests that a preceding voiceless consonant with a following silence is not a sufficient condition for a high vowel to become devoiced — in fact, the correct characterization of this HVD environment aligns with prosodically defined breaks.

References Barr, D., R. Levy, C. Scheepers and H. Tilly. (2013) Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68:255-278. / Kondo, M. (1997) *Mechanisms of vowel devoicing in Japanese*. PhD thesis, University of Edinburgh. / Maekawa, K. and H. Kikuchi. (2005) *Corpus-based analysis of vowel devoicing in spontaneous Japanese: an interim report.* In van de Weijer et al. *Voicing in Japanese*. / Maekawa, K., H. Koiso, S. Furui, and H. Isahara. (2000) Spontaneous speech corpus of Japanese. Proc. 2nd LREC, Athens, 947-952. / Tsuchida, A. (1997) *Phonetics and phonology of Japanese vowel devoicing*. PhD thesis, Cornell University. / Tsuchida, A. (2001) Japanese vowel devoicing: Cases of consecutive devoicing environments. *Journal of East Asian Linguistics*, 10(3):225245. / Varden, J.K. (2010) On vowel devoicing in Japanese. *The MGU Journal of Liberal Arts Studies*, 4(1):223-235. / Varden, J.K. (1998) *On high vowel devoicing in standard modern Japanese: implications for current phonological theory*. PhD thesis, University of Washington.

Production of a non-phonemic contrast by native and non-native speakers: The case of American English flap

Mafuyu Kitahara[1], *Keiichi Taijma*[2], *and Kiyoko Yoneyama*[3]
[1]Waseda University, [2]Hosei University, [3]Daito Bunka University
kitahara@waseda.jp, tajima@hosei.ac.jp, yoneyama@ic.daito.ac.jp

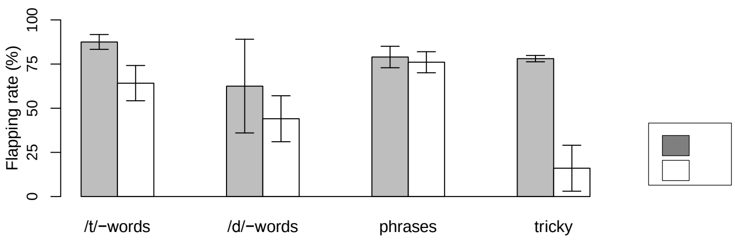
Intervocalic alveolar flapping, the realization of intervocalic /t d/ as an alveolar flap, is commonly observed in American English (AE). While alveolar flaps do not lexically contrast with other sounds in English, flapping is known to occur highly regularly in the speech of native AE speakers. For non-native learners of English, alveolar flaps may not be a critical sound to learn, since they are not contrastive, but it might still be to the learners' advantage to be able to produce them, especially if they want to sound native-like. However, it is unclear how well Japanese learners of English (JE) can produce alveolar flaps in English. It is also unclear how consistently native AE speakers produce alveolar flaps in spontaneous speech. In order to address these questions, the present study conducted a speech corpus analysis and a speech production experiment.

First, native AE speakers' production of alveolar flaps was investigated using the Buckeye Corpus (Pitt *et al.*, 2007). We examined the phonetic transcription of all words, counted those that contained an intervocalic flap, e.g., *letter*, and calculated the flapping rate for each word, i.e., the percentage of tokens of the word that were produced with a flap out of all productions of that word in the corpus. We also counted all cases of across-word flapping in two-word sequences, e.g., *get in*, and calculated the flapping rate for each such "phrase". Results revealed that flapping occurred regularly in many words, as expected. At the same time, however, out of all words that occurred at least five times in the corpus and contained at least one flapped token (N=346), only 24.9% (N=86) showed a flapping rate of 80% or higher. The remaining 75.1% showed lower flapping rates. Mean flapping rate across all 346 words was 48.6%. In addition, many two-word sequences showed across-word flapping. However, mean flapping rate for phrases (N=534) was 38.4% and therefore lower than that for words, and only 12.0% (N=64) showed a flapping rate of 80% or higher. These results suggest that flapping in English is not a categorical phenomenon but is rather a gradient one.

Next, to investigate how often Japanese learners of English produce flapping, we first surveyed a corpus of English sentences spoken by Japanese students from 20 universities across Japan (Minematsu *et al*., 2002). Results revealed that flapping was almost never observed, suggesting that Japanese university students typically do not produce flapping.

Finally, we conducted a preliminary production experiment with five JE learners who had studied at North American universities for at least one semester, in order to investigate the flapping rate of learners with higher English proficiency skills, and to examine the effect of word or phrase context on flapping rate. Flapping rate was calculated for four types of materials (N=50): (1) words with intervocalic /t/, e.g., *writer*, (2) words with intervocalic /d/, e.g., *rider*, (3) phrases with potential across-word flapping, e.g., *get in*, and (4) phrases with "tricky" spelling that may show across-word flapping if the consonant after the alveolar stop is reduced, e.g., *get her*. Figure 1 shows the flapping rates of the five JE learners along with those of AE speakers based on the Buckeye Corpus. Results indicate that flapping rate for the JE speakers was higher than 50% for most words and phrases. Flapping rate was somewhat lower for words with intervocalic /d/, and lowest for phrases with "tricky" spelling.

Figure 1: Flapping rate of words and phrases by AE and JE speakers.



Put together, these results suggest that flapping is widely observed in native AE speakers' conversational speech, although the process appears to be gradient, and not categorical as might be suggested by a strictly rule-based account. Furthermore, while JE learners with no experience living in North America produce little to no flapping, just one semester of experience seems to be sufficient for learners to exhibit some amount of flapping. Nevertheless, more extensive learning is presumably necessary to achieve native-like performance.

References

Minematsu, N., Tomiyama, K., Yoshimoto, K., Shimizu, K., Nakagawa, S., Dantsuji, M., Makino, S. 2002. English Speech Database Read by Japanese Learners for CALL System Development. *Proc. Int. Conf. Language Resources and Evaluation*, 896-903. / Pitt, M.A., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E. Fosler-Lussier, E. 2007. Buckeye Corpus of Conversational Speech (2nd release) [www.buckeyecorpus.osu.edu] Columbus, OH: Dept. of Psychology, Ohio State University (Distributor).

Trajectories of phonetic variability in spontaneous speech on reality TV

Morgan Sonderegger
McGill University
morgan.sonderegger@mcgill.ca

Introduction How a speaker realizes sounds in spontaneous speech varies due to many factors, such as coarticulation and speaking rate. Less is known about what variation as a function of *time* phonetic variables, such as voice onset time or vowel formants, show within individuals. Shifts in phonetic variables have been examined in *phonetic imitation* studies, of shifts over minutes-hours under exposure to the speech of others (Babel, 2011; Nielsen, 2011), and sociolinguistic studies of shifts between time points years apart (e.g. Sankoff & Blondeau, 2007). By contrast, very little is known about how the speech of individuals varies from day to day, over timescales in between. Understanding such daily variability is important for contextualizing phonetic imitation studies, which are often motivated by the hypothesis that sound change results from an *accumulation* of short-term shifts which occur in conversation (Delvaux & Soquet, 2007). If this is the case, individuals' use of a phonetic variable in spontaneous speech should change from day to day. Fig. 1 shows four broad types of time dependence which could obtain: none; by-day variability (BDV) around a static mean; changing mean with no BDV; changing mean with BDV. This study asks, for the first time: which type(s) of day-to-day time dependence do phonetic variables show within individual speakers, over a period of months?

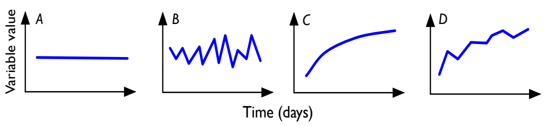
Methods We examine phonetic variation in a British reality television show where speakers live in an isolated house for three months, making it possible to trace the day-to-day dynamics of phonetic variables in contestants' speech. The corpus contains 5 hours of speech from 12 speakers who are on the show for >50 days, and contains only prompted monologues (not conversations), so any observed time dependence is in a variable's "baseline" value beyond short-term shifts which occur in conversation. We consider 5 variables: (1) **VOT** (n = 4858) for word-initial p/t/k, measured automatically (Anonymous, 2014); (2) word-final coronal stop deletion rate (**CSD**; n = 4934), annotated manually; (3)-(5) Lobanov-normalized **F1 and F2 for GOOSE, STRUT, TRAP** vowels (n = 1297, 2351, 1735), measured automatically (Rosenfelder et al., 2012) with hand corrections. For each speaker/variable pair, we fit 12 mixed-effects models corresponding to types of time dependence A-D (allowing for (non)linear time dependence of the variable's mean), controlling for linguistic factors known to affect each variable (e.g., speaking rate, segmental context), and including by-word random effects. We then selected the best model of time dependence for each speaker/variable pair using AIC.

Results Results are shown in Table 1. The clearest pattern is that across variables and speakers, some degree of variability over time (Types B-D) is very common (69% of cases), suggesting that the null hypothesis of stability of phonetic variables in individuals over time can be rejected. (Note that because our model selection procedure will fail to detect variability over time if its magnitude is small enough relative to our sample size, we are *under*estimating the prevalence of B-D.) The prevalence of variability in phonetic variables within individuals over time is consistent with the hypothesis that imitation effects persist on a timescale of hours to days, and hence *could* accumulate. Also consistent with this hypothesis, the magnitude of by-day fluctuations in cases of Types B & D when by-day variability occurs is similar to the effect sizes observed in VOT and vowel formant imitation studies (Babel, 2011; Nielsen, 2011). However, the relatively small proportion of cases which show a changing mean (31% Types C-D) suggest that these day-to-day fluctuations often do not accumulate into longer-term trends, suggesting that additional mechanisms to phonetic imitation may be necessary to explain sound change over the lifespan within individuals.

Table 2: Number of speakers showing each type of time dependence for each variable, with the most common type bolded.

Type	VOT	CSD	GOOSE F1, F2	STRUT F1, F2	TRAP F1, F2
A. None	3	**8**	**3**	1	0
B. By-day variability	4	1	**3**	**5**	2
C. Changing mean	**5**	2	1	1	1
D. Changing mean & BDV	0	0	1	2	**5**

Figure 1: Schematic of possible types of type dependence for a phonetic variable.



References

Anonymous. 2014. Automatic VOT measurement [Computer program] / Babel, M. 2011. Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *J. Phon.* 39, 437-456. / Delvaux, V. & Soquet, A. The influence of ambient speech on adult speech productions through unintentional imitation. *Phonetica* 64, 145-173. Nielsen, K. 2011. Specificity and abstractness of VOT imitation. *J. Phon.* 39, 132-142. / Rosenfelder, I., Fruehwald, J., Evanini, K., & Yuan, J. (2011). *FAVE Program Suite* [Computer program]. / Sankoff, G. & Blondeau, H. 2007. Language change across the lifespan: /r/ in Montreal French. *Lang.* 83, 560–588.

Toward a holistic measure of reduction in spontaneous speech

Michael McAuliffe and Molly Babel
University of British Columbia
michael.mcauliffe@alumni.ubc.ca and molly.babel@ubc.ca

Reduction in spontaneous speech and laboratory speech is a well-studied topic. Traditional methods of quantifying reduction fall broadly into three categories. *Durational approaches* examine variations in the duration of words or vowels (e.g. Bell et al, 2009), but duration does not capture any spectral information. *Spectral approaches* use acoustic measures such as dispersion in the vowel space (e.g. Gahl et al, 2012), but thus far these approaches are limited to a class of sounds, like vowels, and are sensitive to large changes in how words are produced. Finally, *segmental approaches* have quantified reduction in terms of deletion or substitution of segments (e.g. Dilts, 2013), but these approaches are limited by the style of transcription and by gross categorization of spectral changes. This paper presents a single method of quantifying spectral reduction that takes into account the entire word token, is not limited to a specific class of sounds, and can potentially incorporate the information contained in segmental approaches. Using an acoustic similarity metric, we build networks of tokens, and then quantify reduction as the overall similarity of a token to the rest of the network. We predict that more reduced forms will show a higher average similarity to other tokens in the network than more distinct forms.

The data for this study came from the Buckeye Corpus of spontaneous speech (Pitt et al, 2007), which consists of dialogs of 40 speakers, split by gender and age (older, younger) group. To allow for comparisons to the findings in Gahl et al (2012), tokens for this study were CVC content words which were not adjacent to pauses or disfluencies. The similarity metric used for creating the networks follows the one used by Lewandowski (2012) to quantify phonetic convergence. Acoustic waveforms are filtered into eight logarithmically-spaced bands, and each filtered signal is transformed into a normalized amplitude envelope. Given two sets of amplitude envelopes, cross-correlations values for each band are averaged together to get an overall similarity at a given alignment of the envelopes. The similarity value used is the maximum of these values, representing the best possible alignment.

For each speaker in the corpus, tokens matching the Gahl et al (2012) criteria were extracted and connected into a complete, weighted network. For each token in the network, the average similarity value was extracted and then z-score normalized to reduce the influence of the specific networks. A linear mixed-effects model was constructed with Frequency, Neighborhood Density, Speaker Gender, Speaker Age, Previous Speaking Rate, Following Speaking Rate, Previous Conditional Probability, and Following Conditional Probability entered as fixed effects with random intercepts for Speaker and Word and maximal random slopes for Speakers and Words. Numeric predictors were logarithmically transformed and centered. Significant fixed effects were found for Neighborhood Density and Previous Conditional Probability. Words in denser phonological neighborhoods were more similar to other tokens in the phonetic network ($\beta = 0.30$, SE = 0.07, $t = 4.2$), and words with high probability given the preceding word were more similar to other tokens as well ($\beta = 0.41$, SE = 0.17, $t = 2.4$). A linear mixed-effects model for Duration with the same structure replicated the findings in Gahl et al (2012), with significant fixed effects for all factors except for Speaker Age and Gender.

The differences in performance between the model for average similarity and for duration suggest that the average similarity metric should be further refined. The current network creation algorithm is simplistic and compares each token to every other token. Less complete networks with more word tokens across speakers could provide a more reliable estimate for reduction, as the measure is reliant on the network characteristics in addition to the token itself. While the current study limited itself to CVC words for comparison to previous work on the Buckeye Corpus, the method is more generalizable than previous spectral methods, as it uses words as the unit of analysis rather than specific classes of speech sounds. In addition, no extraction of specific features, such as vowel formants, voice-onset time, or spectral center-of-gravity, is needed, nor is it influenced by sub-lexical transcription like segmental or other spectral approaches. Finally, a word-based spectral reduction metric can be applied to corpora that are not annotated at the phone-level, allowing for more corpora to be analyzed.

References

Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. 2009. Predictability effects on durations of content and function words in conversational English. *JML, 60*, 92-111. / Dilts, P. 2013. *Modelling phonetic reduction in a corpus of spoken English using Random Forests and Mixed-Effects Regression*. (Doctoral dissertation). From http://www.ualberta.ca/~pdilts/docs/Dilts_Philip_Fall%202013.pdf. / Gahl, S., Yao, Y., & Johnson, K. 2012. Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *JML, 66*, 789-806. / Lewandowski, N. 2012. *Talent in nonnative phonetic convergence*. (Doctoral dissertation). From http://elib.uni-stuttgart.de/opus/volltexte/2012/7402/. / Pitt, M. A., Dilley, L., Johnson, K., Kiesling, S., Raymond, W., Hume, E., & Fosler-Lussier, E. 2007. Buckeye Corpus of Conversational Speech (2nd release)[www. buckeyecorpus. osu. edu] Columbus, OH: Department of Psychology. *Ohio State University (Distributor)*.

Phonology Constrains the Distribution of the Particle *lah* in Singapore English

James Sneed German[1] & Laurent Prévot[2]

[1]Nanyang Technological University; [2]Aix-Marseille Université & the Laboratoire Parole et Langage

jsgerman@ntu.edu.sg, laurent.prevot@lpl-aix.fr

The meaning of the particle *lah* in Singapore English (SgE) has been characterized as serving a variety of social and discourse functions (see Besemeres & Wierzbicka (2003) for a review). All of these functions, however, reflect a choice on the part of the speaker to convey something. At the same time, *lah* has specific positional and prosodic properties which affect its realization. For example, it occurs only at the ends of utterances and before major prosodic breaks. As *lah* occurs primarily in statements and directives, its realization is governed by the patterns that normally occur in those utterances, such as the presence of a final f0 fall. While some accounts assume distinct tonal variants for *lah* (Wong, 2004), it is never prominence-bearing, and the most common realization involves a falling or sustained low f0 contour.

Utterances differ in terms of the phonological properties of the lexical items that they end with. In particular, they may end with shorter or longer sequences of unstressed syllables ("*He wants to volun**teer**" vs.* "*He's doing ex<u>ercises</u>*"). Since *lah* is unstressed, adding it to the end of an utterance will effectively increase the length of the unstressed sequence. According to principles of eurhythmy (e.g., Hayes, 1984), such long unstressed sequences should be dispreferred. This raises the possibility that constraints on metrical structure may limit the extent to which a speaker can freely deploy *lah* for pragmatic or social functions. In other cases, the phonological context may actually *favor* the presence of *lah*. When the final syllable is stressed, for example, the inclusion of *lah* can relieve tonal crowding by supplying additional segmental material over which to realize the extra tones. If *lah* is sufficiently consistent with the pragmatic context, then the speaker might opt to use it when he or she might not have otherwise.

In sum, there is reason to suppose that the distribution of *lah* is partly determined by the local prosodic context. We tested this hypothesis through analysis of a text-based corpus of conversational spoken Singapore English (ICE-Singapore). We restricted our analysis to the "Private Dialogues (S1A)" in which *lah* is well-represented (1,586 out of 213,555 word tokens). We extracted the last five syllables of each of the 29,855 utterances, of which 942 ended with *lah*. Excluding words not in our lexical database, this yielded 892 contexts with *lah* and 26,952 without. Overall SgE preserves the stress pattern of British English (Bao, 2006), so utterance-final stress patterns were estimated by cross-indexing wordforms with the Celex2 lexical database (Baayen, et al., 1996). As shown in Table 1(left), we categorized tokens according to the number of unstressed syllables preceding the target location (i.e., the utterance boundary or *lah*).

Table 1. *Left*: Frequency of utterance-final contexts with and without *lah* by stress pattern. *Right*: Percentage deviation from the expected value under the null hypothesis.

Stress	No *lah*	*lah*
xxx1	17969	644
xx10	7055	198
x100	1682	44
1000	219	6
0000	27	0

Stress	No *lah*	*lah*
xxx1	-0.3%	+7.9%
xx10	+0.5%	-14.9%
x100	+0.7%	-20.5%
1000	+0.6%	-16.8%
0000	--	--

- "0" : unstressed syllable
- "1": stressed syllable
- "x": syllable of any type or utterance boundary

If the distribution of *lah* is not sensitive to the local prosodic context, then its distribution across categories should reflect the overall proportions across categories. A Chi-square test revealed that this distribution deviates significantly from the expected one ($\chi^2 = 11.77$, $p < 0.01$). ("0000" includes no *lah* tokens and was therefore excluded from the analysis.) Table 1(right) gives the percentage deviation for each cell, showing that the effect is in the expected direction. In other words, *lah* is underrepresented in contexts where its inclusion would have resulted in an unstressed string of two or more syllables, and overrepresented when the final syllable is stressed. These findings support the view that the distribution of *lah* is partly determined by the metrical structure of the local lexical context, and is not solely a matter of speaker intention. These are closely related to findings by Calhoun (2010) and German et al. (2006) showing that phonological constraints can influence accent placement independently of discourse pragmatic considerations. Our findings differ in that *lah* represents a choice of segmental, rather than purely intonational structure. The general pattern is predicted to extend to other particles of SgE as well as similar particles in other languages (e.g., French *quoi*).

References

Baayen, R.H., Piepenbrock, R., & Gulikers, L. (1996). Celex2. *LDC96L14, Linguistic Data Consortium, Philadelphia*. / Bao, Z. (2006). Clash avoidance and metrical opacity in Singapore English. *STUF-Sprachtypologie und Universalienforschung*, *59*(2/2006), 133-147. / Besemeres, M., & Wierzbicka, A. (2003). Pragmatics and cognition: The meaning of the particle *lah* in Singapore English. *Pragmatics & Cognition*, *11*(1), 3-38. / Calhoun, S. (2010). How does informativeness affect prosodic prominence? *Language and Cognitive Processes*, *25*(7-9), 1099-1140. / German, J., Pierrehumbert, J. & Kaufmann, S. (2006). Phonological constraints on nuclear accent placement. *Language 82(1)*, 151-168. / Hayes, B. (1984). The phonology of rhythm in English. *Linguistic Inquiry*, *15*(1), 33-74. / Wong, J. (2004). The particles of Singapore English: A semantic and cultural interpretation. *Journal of Pragmatics*, *36*(4), 739-793.

Word-specific and sub-phonemic representations: yod-dropping and /u/-fronting in Derby

Márton Sóskuthy[1], Paul Foulkes[1], Vincent Hughes[1], Jennifer Hay[2], Bill Haddican[3]

University of York[1], University of Canterbury[2], CUNY–Queens College[3]

marton.soskuthy@york.ac.uk

In this talk, we investigate the relationship between word-level and sub-phonemic representations in speech production using a spontaneous speech corpus. We look at two predictions arising from usage-based models (Pierrehumbert 2002), which both concern interactions between different levels of representation. To test these predictions, we analyse the interaction of yod-dropping and GOOSE-fronting in a cross-generational corpus of Derby English.

The first prediction relates to the coherence of word-level representations across different phonetic environments. In usage-based models, specific lexical items as well as environments can have separate representations. When different environments arise in the same word due to alternations or variation, the coherence of word-level representations may allow phonetic effects to transfer from one environment to the other. Benus & Gafos (2007) demonstrate this type of transfer in Hungarian (a language with V-harmony), where irregular front stems which take back suffixes show backing of the stem (originally due to V-to-V coarticulation) both in suffixed and base forms.

The second prediction is related to the first one: the direction of the transfer of phonetic effects from one sub-phonemic category to the other is predicted to be correlated with their frequencies. The more frequent a variant, the higher its influence will be. Hay & Maclagan (2012) show that this prediction holds in New Zealand English, where the degree of constriction in linking *r* is positively correlated with its relative frequency within specific words (and also specific speakers).

We analyse data from Derby, a city in the north midlands of England. Its dialect is relatively unusual in having both a fronting process affecting GOOSE, and variable yod dropping. Yod has a strong favouring effect on GOOSE fronting, but it is variably lost when it follows the coronal stops [t], [d], [n]. The model above predicts that (i) there will be transfer across the yod-ful/yod-less environments in variable words and (ii) the direction of this transfer will be at least partly determined by the incidence of yod-dropping for specific words.

The data comes from a corpus of 43 speakers, representing four generations born over an 80 year period. This abstract focuses on the speech of the youngest generation. We automatically extracted 11-point F1/ F2 trajectories for over 1500 tokens of the GOOSE vowel (this included the yod as well in yod-ful tokens), and manually corrected them where necessary. The measurements were normalised using the Watt & Fabricius method.

Figure 1 shows *loess* smoothers for (i) yod-ful/yod-less tokens of variably yod-dropping words (NEW), (ii) tokens with preceding coronals but no yod (NOODLE) and (iii) yod-ful tokens where yod-dropping is not possible (CUBE). A comparison of yod-less NEW and NOODLE shows that the fronting effect of the yod transfers to yod-less tokens. Moreover, yod-ful NEW is lower than CUBE, suggesting that the transfer also occurs in the other direction. These observations are supported by linear mixed effects models, and are in line with the first prediction above.

Figure 2 plots predictions from a linear mixed effect model, which was run on tokens from the NEW class, and included the proportions of yod-ful variants for specific words as a predictor, which emerged as significant. The difference between the dotted (low yod proportion) / solid lines (high yod proportion) shows that words with more yod-ful variants have generally higher F2, which confirms the second prediction above.

These results provide support for usage-based models of speech production, where word-level and sub-phonemic representations both have an influence on production. Analysis of the full corpus is ongoing to test the robustness of these effects, and see how they interact with the gradual process of change that has affected the GOOSE lexical set.

Figure 1

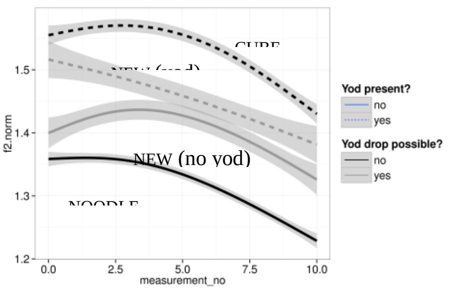
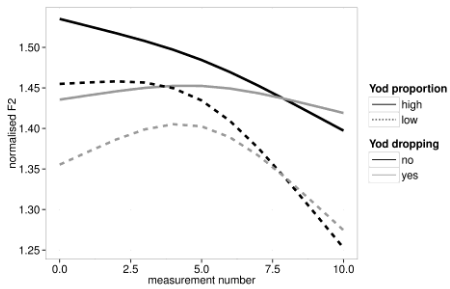


Figure 2



References

Hay, J., Maclagan, M. 2012. /r/ sandhi in early 20th century New Zealand English. Linguistics 50:4, 745–763. / Pierrehumbert, J. 2002. Word-specific phonetics. In: Gussenhoven, C. and Warner, N. (Eds.) *Laboratory Phonology VII*. Berlin: Mouton de Gruyter. / Benus, S., Gafos, A. 2007. Articulatory characteristics of Hungarian 'transparent' vowels. Journal of Phonetics, 35, 271-300.

Verbal feedback: positioning and acoustics of French "ouais" and "oui"

Laurent Prévot, Jan Gorisch
Aix Marseille Université & CNRS, Laboratoire Parole et Langage, France
laurent.prevot@lpl-aix.fr jan.gorisch@lpl-aix.fr

Embedded in a larger study that examines the form-function relationship for French positive feedback markers this study examines the form of *'oui'* and *'ouais'* with respect to their relative placement to the own speaker's and the interlocutor's speech production, and with respect to acoustic features including F0, aperiodicity and intensity.

Previous corpus-based research on these items has shown the importance of positional features for determining the choice of a specific item (Gravano et al., 2012). Neiberg et al., (2013) has shown the importance of phonological operations in Swedish (similar to our *'oui'* vs. *'ouais'* opposition) for communicating different attitudes. Other studies took a qualitative approach towards the phonetics of response tokens in naturally occurring talk in interaction. Gardner (2001) shows that the response token *'mm'* in English has specific acoustic features (fall-rising, falling or rise-falling pitch) when it is used in specific sequences of talk (continuation, acknowledgement, assessment). Ogden (2010) found that specific phonetic-prosodic properties (loudness, F0 height, F0 span and voice quality) of complaints have specific interactional consequences. In French, *'oui'* vs. *'ouais'* is assumed to be opposed in terms of level of language ('*ouais'* being more colloquial). However, their simultaneous presence in most interactive situations as well as some recent descriptive work (Péroz, 2009) tend to suggest different communicative functions for these two items. Here, we apply basic machine-learning techniques on a large dataset of these feedback items. The driving research question is how the various positional and acoustic properties interact and produce distinct properties of feedback items. The hypothesis is that independent from the lexical content (*'ouais'* vs. *'oui'*), the selected features can be used to train a classifier that distinguishes between different speech-exchange situations, e.g. conversational vs. task-oriented dialogue.

The material used for this study contains two corpora. The CID corpus (Bertrand et al., 2008) are face-to-face conversations that have the mere instruction to talk about "particular" events. The Aix Maptask Corpus (Gorisch et al., 2014), which is the French version of the famous Edinburgh corpus. Two conditions are recorded: face-to-face vs. remote interaction.

Since manual coding for such short items has been found to be difficult (D'Imperio et al., 2013) for the time being, we used a completely automatic and data-driven approach using only the sound files and the manual transcription that has been force aligned using SPPAS software (Bigi, 2012). Our final dataset was built from the 12 hours (CID: 8 hours, Maptask: 4 hours) of a transcribed and force-aligned corpus. We extracted from this data sets all the isolated *'oui'* (N: 352) and *'ouais'* (N: 1359) (isolated = item surrounded by over 200 ms pauses). We then extracted a set of features concerning their environment (including normalized duration of overlap, normalized position in the dialogue, position with regard to the other speaker's speech) and more acoustic features (duration, intensity median, F0 slope, F0 standard deviation, aperiodicity, …). We were then in the position of classification exploratory experiments. We used Weka (Hall et al., 2009) to try to predict meta-data information (*'oui'* vs. *'ouais'*, corpus, speaker role) based on our feature set. About *'oui'* vs. *'ouais'* classification results, Naïve Bayes and Simple Logistic classifiers achieve about 79% accuracy without having speaker. This is about the score of majority-class baseline. When speaker information is added, this score rises to 88%. However, inspection of the decision-tree (J48) version of the classification shows that the increase in accuracy is partly due to a majority-class choice per speaker. Other features used by the classifiers include overlap information, duration and F0 standard deviation. The results on speaker role prediction (Co-narrator for the CID, Giver and Follower for the MapTasks) are stronger with an accuracy of 79.5% (without using meta-data) using first F0 slope, then duration, overlap features and intensity. Finally, corpus prediction, gave 71.5 % accuracy using mostly duration and then overlap and slope features. The current work consists of refining and improving the features. We are also finishing to transcribe a third corpus of 2 hours argumentation and negociation.

References

Bigi, B. (2012). SPPAS: a tool for the phonetic segmentations of speech, *8th LREC*, Istanbul. / D'Imperio, M., Petrone, C., Prévot, L. (2013). Basic prosodic transcription of short French feedback utterances. In *Workshop Advancing Prosodic Transcription for Spoken Language Science and Technology II*. / Bertrand, R., Blache, P., Espesser, R., Ferré, G., Meunier, C., Priego-Valverde, B., Rauzy, S. (2008). CID — Corpus of Interactional Data — Annotation et exploitation multimodale de parole conversationnelle. *Traitement Automatique des Langues 49(3)*, 1-30. / Gardner, R. (2001). *When listeners talk: Response tokens and listener stance*. John Benjamins. / Gorisch, J., Astésano, C., Gurman Bard, E., Bigi, B., Prévot L. (2014). Aix Map Task corpus : The French multimodal corpus of task-oriented dialogue, *9th LREC 2014*, Reykjavik. / Gravano, A., Hirschberg, J., Benus, S. (2012). Affirmative cue words in task-oriented dialogue. *Computational Linguistics 38(1)*, 1-39. / Hall, M., Eibe, F., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I. (2009). The WEKA Data Mining Software: An Update; *SIGKDD Explorations, 11(1)*. / Neiberg, D., Salvi, G., Gustafson, J. (2013). Semi-supervised methods for exploring the acoustics of simple productive feedback. *Speech Comm. 55*, 451-469. / Ogden, R. (2010). Prosodic constructions in making complaints. In D. Barth-Weingarten, E. Reber, & M. Selting, *Prosody in Interaction* (pp. 81-103). John Benjamins. / Péroz, P. (2009). « « On ne dit pas ouais ! » », *Langue française* 1/2009 (n° 161), 115-134.

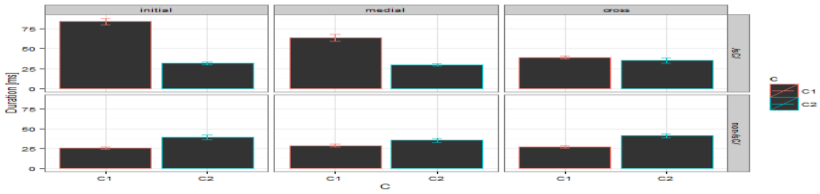
Dissecting the Consonant Duration Ratio

Erika Pillmeier, Stavroula Sotiropoulou and Adamantios Gafos
Department of Linguistics, University of Potsdam
erika.pillmeier@gmail.com, stavroula.sot@gmail.com,
adamantios.gafos@uni-potsdam.de

It has been argued that the duration ratio (DR) of the two consonants in C1C2 serves as a diagnostic of syllabification: if greater than 1, then the cluster C1C2 is parsed as a (complex) syllable onset; if (approx.) equal to 1, the cluster is a coda (C1) – onset (C2) sequence, split across two syllables (Christie, 1977; Redford, 2003; Redford and Randall, 2005). If valid, this diagnostic would provide a straightforward way of assessing syllabic organization. We **assessed the validity of the DR** using the X-ray microbeam corpus (Westbury, 1995), which due to its size (57 speakers, Am. English, kinematics & acoustics) and variety of data (isolated words, phrases, but also whole paragraphs) permits a robust test. We also informed our assessment by another archive of Electromagnetic Articulometry (EMA) and acoustics, which allows finer control on comparisons but more limited data. Our corpora allow us to also extend an evaluation of the DR to word-medial clusters, in addition to word-initial and cross-word contexts wherein the DR has been studied before.

For the purposes of this abstract, we refer to articulatory data; however, at the meeting, we will report also acoustic results. From the first corpus, DRs for /sn sm fl gl tw/ were extracted from 24 to 35 speakers per cluster. From the second corpus, DRs from /sp sk br kl/ were taken from 2 speakers. One key result can be seen by comparing two cluster sets, fricative-stop (Fric-C) and stop-liquid (Stop-L) in Figure 1. For Fric-C clusters, the DR shows the expected pattern. The top row of Figure 1 shows why this is so: in Fric-C, while C1 expands significantly from cross word to medial to initial position ($p < 0.001$ and $p < 0.03$, respectively), C2 remains stable. Any change in DR is thus the result of C1's length modulation. For Stop-L, however, C1 and C2 are stable across the three positions, with no (sign.) difference between any of the three for either consonant. Thus, Stop-L does not show the pattern expected (all positions: DR < 1).

Figure 1: Plateau durations per consonant class and position in word (Top row: Fric-Stop; Bottom row: Stop-L).



Whence these patterns? Independently known prosodic effects seem implicated. Briefly, in cross-word Fric-Stops, the word-final Fric undergoes final lengthening and the stop initial lengthening (see Byrd and Choi, 2010 and references therein); as a result their DR is approximately 1. In word-initial clusters, Fric undergoes initial lengthening, and in word-medial clusters it undergoes stress-related lengthening (in keeping with previous work, for all word-medial Fric-Stops, stress is on the vowel following the cluster). As a result, in the two latter cases, the DR is greater than 1. These patterns do not hold for clusters with a stop as a first member, because stops are not subject to prosody-driven effects (final, initial, and stress-related lengthening) to the same degree as fricatives (Oller, 1973). In sum, the DR is not an index to syllabification, but rather a byproduct of prosodic modulations. The results in Fig. 1 show only a subset of the patterns in our corpora and should not be taken to mean that the DR expected pattern is found for Fric-Stop but not for Stop-L: when the convergence of prosodic conditions leading to the expected patterning of the DR are not met, even Fric-Stops fail to show the pattern. For example, according to the expected pattern, DRs between word-initial and cross-word /sC/ should be different with the former DR greater than the second. However, word-initial /sn/ taken from a *medial utterance position* and cross-word /sn/ show no significant difference in DR ($p=0.13$). The lengthening of /s/ in that word-initial position is not as strong as would be required to raise its DR to a value different from the cross-word condition. Indeed, across our corpus in addition to main effects of C1 manner ($F(1, 3)=30.2$, $p=.0119$) and word position ($F(1, 3)=12.337$, $p=.0391$), we also find an interaction between utterance position and word position ($F(1, 3)=18.963$, $p=.0224$).

In sum, duration ratio patterns previously thought to be related to syllabification derive from levels of representation or prosody which are not specifically syllabic. These patterns are conditioned by prosodic effects and manner contrasts.

References

Byrd, D. & Choi, S. 2010. At the juncture of prosody, phonology, and phonetics: the interaction of phrasal and syllable structure in shaping the timing of consonant gestures. *Laboratory Phonology 10*. Mouton de Gruyter. / Christie, W.M. 1977. Some multiple cues for juncture in English. *General Linguistics* 17, 212-222. / Oller, D. K. 1973. The effect of position in utterance on speech segment duration in English. *J. Acoust. Soc. Am*. 54, 1235-47. / Redford, M.A. 2003. Cognitive template for a phonetic correlate of syllable structure. In: M. J. Solé, D. Recasens & J. Romero (eds.), *ICPhS* 15, 2261-2264. Casual Productions. / Redford, M. A. & Randall, P. 2005. The role of juncture cues and phonological knowledge in English syllabification judgments. *Journal of Phonetics* 33, 27-46. / Westbury, J. R. 1994. *X-ray microbeam speech production database user's handbook*. University of Wisconsin-Madison.

Derived onsets in Spanish: an experimental study of resyllabification

Patrycja Strycharczuk & Martin Kohlberger

CASL, Queen Margaret University, Leiden University

pstrycharczuk@gmail.com, m.kohlberger@hum.ledeinuniv.nl

Word-final consonants in Spanish are commonly assumed to undergo resyllabification across a word boundary before a following vowel, e.g. /palmas#altas/ 'tall palm trees' is realized as [pal.ma.s|al.tas]. However, in many dialects of Spanish, word-final prevocalic consonants ('derived onsets') pattern phonologically with canonical codas and distinctly from canonical onsets. For instance, /s/ undergoes voicing in codas and in derived onsets in Quito Spanish, but not in canonical onsets (Robinson, 1979). A similar prosodic distribution is reported for /s/-aspiration in Caribbean Spanish and Rio Negro Argentinian Spanish (Kaisse, 1999). This property of derived onsets has been subject of much interest in the phonological literature, and has led some linguists to question whether resyllabification indeed applies in all Spanish dialects (Robinson, 2012). As much of the debate concerning resyllabification is based on impressionistic rather than instrumental data, the present study sets out to evaluate the status of Spanish resyllabification using experimental evidence.

In our investigation, we focus on /s/ in Castilian Spanish, where /s/ resyllabification does not interact with any phonological segmental processes. We recorded 10 native speakers of Castilian Spanish read 4 repetitions of 30 test items which included /s/ in the derived onset position (e.g. *peces asados* 'fried fish'), as well in word-medial onsets (*gran pesadilla* 'big nightmare'), word-initial onsets (*cruce sagrado* 'sacred crossing'), word-medial codas (*diez estatutos* 'ten statues'), and word-final codas (*viajes pagados* 'paid trips'). The test items were embedded in a fixed carrier phrase. The size of the word, stress pattern and the segmental context were controlled for. We measured /s/-duration and the duration of the flanking segments, based on manual segmentation of acoustic data in Praat. We analysed the duration measurements using linear mixed-effects regression with random intercepts for speaker and item. The *p*-values were calculated using Satterthwaite's approximation.

Our results show that derived onset /s/ is on average 18ms longer than /s/ in the canonical coda position ($p<.001$). At the same time, however, derived onset /s/ is also distinct from onset /s/: it is on average 12ms shorter than /s/ in word-initial onsets ($p<.001$), and 7ms shorter than /s/ in word-medial onsets ($p=.05$). Thus, derived onset /s/ shows intermediate duration effects compared to canonical codas and onsets. We also note that the distribution of /s/ duration in derived onsets /s/ is unimodal, which indicates that derived onsets form a single category, and that their intermediate duration is not a result of the resyllabification rule applying optionally.

The intermediate duration of /s/ in the derived onset position suggests that derived onsets are prosodically distinct from both canonical codas and canonical onsets. We interpret this result as evidence for incomplete resyllabification of word-final prevocalic /s/: resyllabification applies, but the resyllabified consonant nevertheless retains partial similarity to coda /s/, or more specifically word-final /s/. These results are consistent with previous findings on French (Fougeron, 2007) and English (Gick, 2003), which also point towards an intermediate prosodic status of derived onsets. Thus, accumulating experimental evidence from different languages suggests that resyllabification across word boundaries may tend to create unique prosodic categories which share phonetic features of canonical onsets and canonical codas. We propose that the representational tension created by such partial similarity effects may play a role in sound change, sometimes conditioning derived onsets to pattern with codas, as seen for instance in Quito Spanish /s/-voicing, or /s/-aspiration in Caribbean or Rio Negro Argentinian Spanish. We further consider the potential role of duration in the actuation of these changes, and discuss to what extent the prosodic asymmetries we see in the diachronic development may be conditioned by increased duration in canonical onsets compared to other prosodic environments.

References

Fougeron. C. 2007. Word boundaries and contrast neutralization in the case of *enchaînement* in French. In: Cole, J., Hualde, J.I. (eds.), *Papers in Laboratory Phonology IX: Change in phonology.* Berlin: Mouton de Gruyter, 609-642. / Gick, B. 2003. Articulatory correlates of ambisyllabicity in English glides and liquids. In: Local, J., Ogden, R., Temple, R. (eds.), *Papers in Laboratory Phonology VI: Constraints on phonetic interpretation.* Cambridge: CUP. 222-236. / Kaisse, E. 1999. Resyllabification precedes all segmental rules. In: Authier, J.-M., Bullock, B. E., Reed, L. A. (eds.), *Formal Perspectives on Romance Linguistics: Selected Papers from the 28th Linguistic Symposium on Romance Languages (LSRL XVIII),* Amsterdam; Philadelphia: John Benjamins, 197-210. /Robinson, K. L. 1979. On the voicing of intervocalic *s* in the Ecuadorian Highlands. *Romance Philology* 33, 132-143. / Robinson, K. L. 2012. The dialectology of syllabification: A review of variation in the Ecuadorian Highlands. *Romance Philology* 66, 115-145.

Tashlhiyt syllabification: Perceptual evidence

Rachid Ridouane[1] *& Pierre Hallé*[1]

[1] Laboratoire de Phonétique et Phonologie (CNRS/Sorbonne Nouvelle)

{rachid.ridouane; pierre.hallé@univ-paris3.fr}

Syllabification in Tashlhiyt is claimed to be sensitive to two basic constraints, a sonority-based constraint that hierarchizes the types of segments that may be syllable nuclei and margin constraints that regulate syllable boundaries (Dell & Elmedlaoui 2002, Prince & Smolensky 1993/2004). Among the margin constraints, the one prohibiting complex onsets is highly active and can never be violated. A consequence of this is that all prevocalic consonant clusters are parsed as heterosyllabic. In this study we are interested in assessing whether the heterosyllabicity of such consonant sequences is reflected at the perceptual level and provide metalinguistic judgments about how they are segmented into syllables by native speakers/listeners. We conducted two experiments: a part-repetition task to test how native speakers syllabify intervocalic two-consonant clusters, and an explicit segmentation task asking native speakers to report the different "parts" of the words they were presented with. Twenty native speakers (mean age 24), all students at the university of Agadir, participated in both experiments. In addition ten illiterate native speakers (mean age 45) participated to the second experiment. In the remaining of this abstract, the method and results of the first experiment will be detailed. Results from the second experiment will be only briefly outlined.

A set of 56 (C)<u>VCCV</u>(C) items, included within a larger corpus with different word structures, were used for experiment 1. The items varied in terms of morphological structure and sonority profile of the CC sequence. Twenty-nine items were monomorphemic and 27 were heteromorphemic. We controlled morphological structure in order to assess whether native speakers' segmentation of utterances corresponds to morpheme boundaries: for instance, /i-sli/ '3ms-touch' segmented in the two parts [i] and [sli]. The sonority profile of the CC sequence defined three groups of items: rising, falling, and plateau profile. The 56 items were coded separately for the first part and the second part conditions. For the first part condition, there were three possible types of response: (C)V, (C)VC, and (C)VCC (e.g., [isli]: i, is, isl). For the second part condition, there were also three possible types of response: V(C), CV(C) and CCV(C) (e.g., [isli]: i, li, sli).

Principles governing the syllabification of consonant sequences such as the Onset Maximization Principle (Clements & Keyser 1983) and the Sonority Sequencing Principle and its coda-to-onset qualification (Clements 1990), have been claimed to play a fundamental role in determining possible onsets. These two principles suggest a cross-linguistic preference for V.CCV over VC.CV syllabifications as long as CC has a rising sonority profile. This preference seems universal in the case of obstruent plus liquid (OL) clusters. If the claim that Tashlhiyt bans complex onsets is correct, we expect VCCV items to be parsed as VC.CV, regardless of the sonority profile of the CC cluster, and whether or not it is an OL cluster.

Results of the part-repetition task provide clear evidence that Tashlhiyt native speakers uniformly analyze intervocalic CC clusters as heterosyllabic. For Part 2, participants overwhelmingly preferred CV(C) over other responses (94%). For part 1, they overwhelmingly preferred (C)VC over other responses (97%). The morphological structure and sonority profile of the sequences did not significantly influence listeners' judgments. Our pattern of results is thus in agreement with the phonological analysis that systematically parses such clusters as heterosyllabic. Strikingly, this pattern held for OL clusters as well as for other clusters, although OL clusters are tautosyllabic in most languages. Results from experiment 2 provide additional evidence that complex onsets are dispreferred. In test words with a #CCV(C) or #CCCV(C) structure, more than 94% of the responses corresponded to a bisyllabic parsing. Consider for example two such items: [gli] 'guide' and [tχwa] 'it is empty'. These forms sound like English *glee* and French *trois*, respectively. But whereas they are monosyllabic in English and French, Tashlhiyt native speakers overwhelmingly judged them as having two parts: [g] and [li] for the former, and [tχ] and [wa] for the latter.

References

Dell, F. Elmedlaoui, M. 2002. Syllables in Tashlhiyt Berber and in Moroccan Arabic. Dordrecht: Kluwer. / Prince, A., Smolensky, P. 1993/2004. Optimality Theory: constraint interaction in generative grammar. Malden: Blackwell Publishing. / Clements, G.N. 1990. The role of the sonority cycle in core syllabification, in: Kingston, J., Beckman, M. (eds.), Papers in laboratory phonology I. Cambridge: Cambridge University Press, 283–333. / Clements, G.N., Keyser, J. 1983. CV phonology: a generative theory of the syllable. Cambridge: MIT Press.

Listener-specific perception of speaker-specific productions in intonation

Francesco Cangemi, Martine Grice
IfL Phonetik, University of Cologne
{fcangemi,martine.grice}@uni-koeln.de

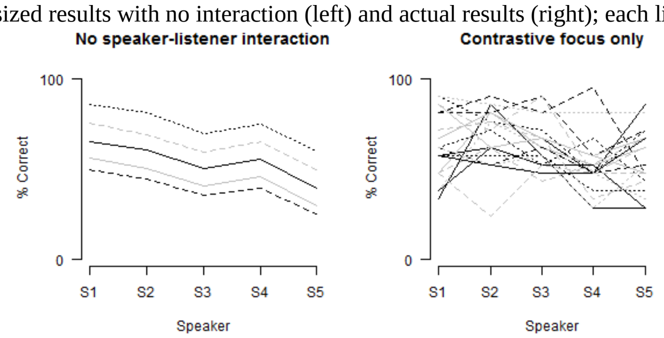
§1. Individual-specific behaviours for phonological contrasts at the segmental level have long been documented in production and perception. The interaction between the two – that is, the hypothesis that an individual's identification or discrimination performances are related to her articulation strategies or acoustic output – has been explored with the aim of supporting frameworks which posit a link between production and perception, yielding contrasting results (Bell-Berti et al. 1979, contra Paliwal et al. 1983). More recently, the issue has been explored in relation to speaker intelligibility (Newman et al. 2001, Hazan et al. 2013), and as a source of support for models of speech production in which articulatory movements are planned primarily in auditory space (Perkell et al., 2004).

§2. Beyond the segmental level, considerably less research has been devoted to speaker- and listener-specific behaviours in intonation (but see for production Niebuhr et al., 2011), and none to the interaction between the two. In this contribution, we document such an interaction in the encoding and decoding of pitch accent categories in German. In the production study, articulatory and acoustic recordings from 5 speakers uttering 7 repetitions of 4 target sentences in 4 dialogue contexts eliciting different focus interpretations (out of focus, broad, narrow and corrective focus) were collected (560 items). Speakers were found to vary with respect to the use of different acoustic cues (f0 peak alignment, f0 range, duration of target word) across the four focus conditions. In the perception study, 20 subjects provided forced-choice identifications through context-matching of 420 utterances from the production study (8400 items), evenly distributed across speakers, sentences and focus conditions.

§3. As could be expected, results pooled across focus conditions and examined through likelihood ratio tests on mixed effects models show that (i) some speakers are perceived better than others and (ii) some listeners perceive better than others. The main finding is that (iii) some listeners perform better on particular speakers, whereas other listeners perform badly on these speakers, but better on other speakers. Thus, results reveal a clear interaction between speaker-specific and listener-specific behaviours (Fig. 1 for one of the four focus conditions). Even more interestingly, (iv) the interaction between speaker-specific and listener-specific effects is modulated by focus condition, e.g. corrective focus items show considerably fewer listener-speaker interaction effects than broad focus items.

§4. These findings not only extend the available knowledge on production-perception links and speaker intelligibility beyond the segmental level (see §1), they are also particularly relevant to the issue of categories in intonation. Building on dynamic modelling (e.g. Gafos and Benus, 2006) of the relations between phonetics and phonology (e.g. Pierrehumbert et al., 2000), recent approaches to categoriality in intonation have stepped away from a monothetic perspective and suggested a distributional view (Nava, 2010). In this perspective, phonetic realizations of phonological categories are distributed along a continuum. Our finding that the strength of speaker-listener interactions depends on the intended focus structure (§3.iv) might help explain the differences in shape of distributions for the various categories, pointing to the existence of more compact vs. more diffuse categories in intonation.

Figure 1: Hypothesized results with no interaction (left) and actual results (right); each line represents one listener



References

Bell-Berti, F. et al. 1979. Some relationships between speech production and perception. *Phonetica* 36, 373-383. / Gafos, A., Benus, S. 2006. Dynamics of phonological cognition. *Cognitive Science* 30(5), 905-943. / Hazan, V. et al. 2013. The impact of variation in phoneme category structure on consonant intelligibility. Proc. ICA Montreal 19, 1-6. / Nava, E. 2010. *Connecting phrasal and rhythmic events: Evidence from second language speech*. PhD thesis, USC. / Newman, R.S. et al. 2001. Perceptual consequences of within-talker variability in fricative production. *JASA* 109, 1181-1196. / Niebuhr, O. et al. 2011. Are there "shapers" and "aligners"? Individual differences in signaling pitch accent category. *Proc. 17th ICPhS*, 120-123. / Paliwal, K.K. et al. 1983. Correlation between production and perception of English vowels. *J. Phon* 11, 77-83. / Perkell, J.S. et al. 2004. The distinctness of speakers' productions of vowel constrasts is related to their discrimination of the contrasts. *JASA* 116, 2338-2344 / Pierrehumbert, J. et al. 2000. Conceptual foundations of phonology as a laboratory science. In: Burton-Roberts, N. et al. (eds), *Phonological knowledge*. OUP, 273-303.

Focus Intonation in Turkish

Canan Ipek[1]*, Sun-Ah Jun*[2]

[1]Department of Linguistics, University of Southern California, Los Angeles

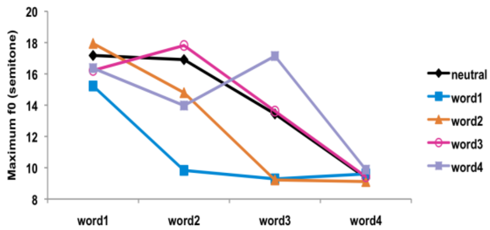
[2]Department of Linguistics, University of California Los Angeles, Los Angeles

ipek@usc.edu, jun@humnet.ucla.edu

This study investigates the intonational properties of focus in Turkish both phonologically and phonetically. In stress-accent languages such as English, focus is phonologically marked by a certain type of pitch-accent on the focused word and deaccenting all post-focal words (e.g., Beckman and Pierrehumbert 1986, Jun 2014). Phonetically, a focused word is usually realized with expanded pitch-range while post-focal words are in compressed pitch range (e.g. Botinis et. al. 1999, Rump & Collier 1996, Xu & Xu 2005). However, Turkish seems to be an exception to this generalization. Ipek (2011) has shown that in Turkish on-focus high tone is substantially compressed in pitch-range while the pitch range of the syllable immediately preceding the focused word is substantially increased. In this paper, we will examine the tonal properties of focus in more detail by adopting the phonological model of Turkish intonation (Ipek & Jun 2013), where prenuclear words are analyzed to carry H* on their stressed syllable and the nuclear word to carry !H*. In that model, the last syllable of a word immediately preceding the nuclear word is analyzed to have a High phrasal tone.

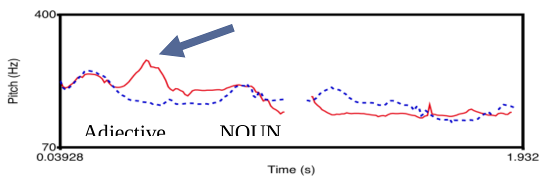
Two groups of declarative sentences were designed to study focus realization. The first group contained five four-word sentences where each word had word-final stress. Sentences were produced by five speakers under five focus conditions: neutral focus and narrow focus on each word (triggered by wh-questions). Pitch tracks were analyzed and labeled phonologically and the maximum f0 values for each word and syllable in on-focus, pre-focus, and post-focus domain were compared with those in their neutral counterpart. We found that a narrowly focused word in Turkish is marked by a !H* pitch accent on its stressed syllable when the word is not phrase-initial (thus similar to the nuclear accent in the neutral focus condition), but by H* when it's phrase initial (even though the H* is not as high as that in the neutral counterpart). In addition, all pitch accents following the focused word were deleted, showing low flat f0. Furthermore, the f0 of pre-focus words was lower than that of the neutral counterpart except for the f0 of word immediately before the focused word, which was substantially higher than its neutral counterpart, suggesting that this high f0 is triggered by the following focus. See Figure 1 for comparison:

Figure 2: Max f0 value (in semitone) for each word, realized on the word-final stressed syllable, under five focus conditions.



However, since the high toned syllable just before the focused word is stressed, receiving H*, one might argue that the High tone is not triggered by the following focus. If focus indeed triggers a high tone aligned with the right-edge of the immediately preceding word, we would expect to see a high tone on a non-finally stressed word when it immediately precedes a focused word. To test this hypothesis, we created a second group of sentences which contained a non-finally stressed word in phrase-medial position (e.g., Adj in "Adj+N" NP, N1 in "N1-GENITIVE + N2" NP). We chose phrase-medial words as the target because it is known that in Turkish phrase-final words carry a High tone on their last syllable regardless of their stress location. Therefore, it is expected that in the sentences we created, the first word in the subject NP will not have a word final high tone when the sentence is produced in a neutral focus condition, but will have a high tone when the following word is narrowly focused. Our findings confirm the prediction (see Fig.2). In sum, narrow focus in Turkish is realized by compressed pitch range on the focused word, phonologically carrying !H*, as well as by expanded pitch range on the immediately pre-focus word, marking the left edge of a focused word. That is, Turkish is similar to other languages in manipulating pitch range to cue focus but differ from them in tone-text alignment.

Figure 3: F0 track of '*Bembeyaz duvarlar haftaya maviye boyanacak* 'Very white walls will be painted in blue next week', showing an f0 peak at the end of Adj. when the N is focused (red solid line) vs. no f0 peak in the neutral focus condition (blue dotted line).

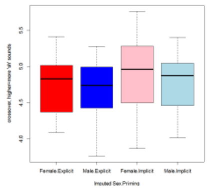


Explicit and Implicit Gender Priming in Fricative Perception

Benjamin Munson
University of Minnesota, Minneapolis, USA
Munso005@umn.edu

The identification of sounds on an /s/ to /ʃ/ continuum differs when they are combined with vowels produced by men or by women (Mann & Repp, 1980). Gender effects on fricative perception can occur even when there are no clear acoustic cues to gender. Strand and Johnson (1996) demonstrated that fricatives are labeled differently depending on whether they are played concurrent with male or female faces, at least when combined with vowels whose acoustic characteristics are gender ambiguous. This *gender normalization effect* has recently been replicated by Munson (2011) and by Winn, Rhone, Chatterjee, and Idsardi (2013). A potential weakness of the methods used previously is that they impute speaker gender explicitly, by using a clearly gendered voice or by pairing auditory stimuli with a clearly gendered face. Both of these methods might result in study participants invoking stereotypes about how men and women's speech differ. Consequently, it is not clear whether participants' perceptual adjustments in those studies reflect authentic knowledge of gender differences in speech (as proposed by Strand and Johnson, 1996), or stereotypes that are invoked (or perhaps suppressed) when gender is mentioned explicitly. It is methodologically feasible to *imply* social categories like talker gender more subtly (e.g., Hay and Drager, 2010). The current study contrasted an explicit method for suggesting talker gender (showing pictures of purported male or female talkers) to an implicit one (listening to sentences whose content was either stereotypically female or stereotypically male prior to identifying fricatives). We examined whether gender normalization effects in fricative perception are equally robust in these two conditions. A finding that there are different fricative identification patterns in the two conditions will help us better understand the cognitive processes that are involved in gender normalization. Listeners participated in one of four conditions. Each condition had a *prime* and a *test* phase. In the implicit condition, the prime phase consisted of listening to sentences that suggested either female (*implicit-female*, n=20) or male (*implicit-male*, n=20) content, and judging their grammaticality. The extent to which each sentence suggested a man or a woman was determined by a pilot group of 10 raters. Recorded sentences were modified acoustically so that their f0 and resonant frequencies were ambiguous between those for men and women, as judged by pilot participants. In the explicit condition, the prime phase consisted of listening to the same speaker producing grammatical and ungrammatical sentences whose content was not clearly gendered, as determined by a pilot group. The test phase involved labeling a series of stimuli as either *sack* or *shack*. The stimuli were created by combining a nine-step synthetic /s/-/ʃ/ series with a production of /æk/ by the speaker from the prime phase, acoustically modified to be gender neutral. The explicit group did this task while viewing either a man's face (*explicit-male*, n=20) or a woman's face (*explicit-female*, n=20). Logit mixed-effects regression examined the effect of stimulus step, priming condition (implicit vs. explicit), and imputed talker gender on fricative categorization. The interaction between priming condition and talker gender was significant. In a model using only the participants in the implicit condition, the effect of gender was not significant. The effect of gender was significant in a model using only the participants from the explicit condition, and it was in the expected direction: more *sack* judgments were elicited for trials where a man's face was used, as shown in Figure 1. Together, these data show that gender normalization effects in fricative perception occur only when the talker's gender is suggested very explicitly by showing a picture of the talker. This finding questions an interpretation of Strand and Johnson's effect as evidence that social variables influence speech perception pervasively.

Figure 4. Crossover points from individual participants.



References

Hay, J., Drager, K. 2010. Stuffed toys and speech perception. *Linguistics, 48*, 865-892 / Mann, V. A., Repp, B. H. 1980. Influence of vocalic context on the perception of [ʃ]-s] distinction: I. Temporal factors. *Percept. Pyshophys. 28,* 213-228 / Munson, B. 2011. The influence of actual and imputed talker gender on fricative perception, revisited. *J. Acoust. Soc. Am, 5,* 2631-2634 / Strand, E., Johnson, K. 1996. Gradient and visual speaker normalization in the perception of fricatives. In *Natural Language Processing and Speech Technology: Results of the 3rd KONVENS Conference, Bielfelt*, ed. D. Gibbon, 14-26 / Winn, M., Rhone, A., Chatterjee, M., Idsardi, W. 2013. The use of auditory and visual context in speech perception by listeners with normal hearing and listeners with cochlear implants. *Frontiers in Psychology, 4,* 824

Articulatory Gestures and Assimilation of Mora Nasal /ɴ/ in Japanese

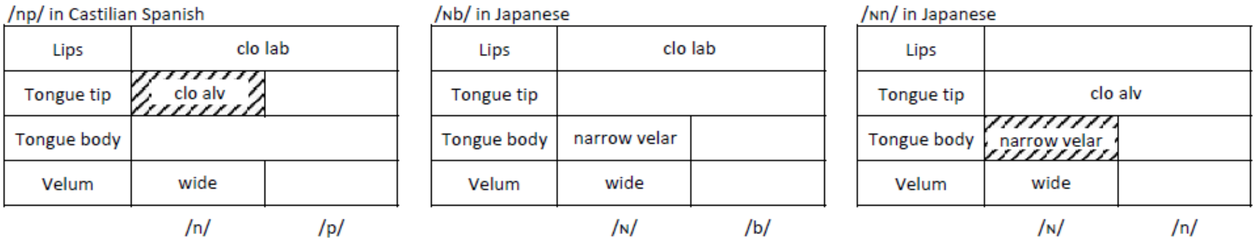
Ai Mizoguchi[1] *and D. H. Whalen*[1, 2]
[1]The Graduate Center, City University of New York
[2]Haskins Laboratories
amizoguchi@gc.cuny.edu, dwhalen@gc.cuny.edu

Gestural behaviors in assimilation can be explained in two ways: 1) gradient and continuous or 2) categorical and discrete. Articulatory Phonology (AP; Browman & Goldstein, 1989; Browman & Goldstein, 1990) utilizes the former. According to the theory of AP, the gestures in assimilation blend, slide and are never deleted or added. Some evidence supporting this AP explanation has been provided by the revelation of hidden gestures for assimilated unperceived phonemes (Browman & Goldstein, 1990). However, evidence for categorical assimilation has also been shown in an investigation of Castilian Spanish nasal assimilation (Honorof, 1999). This result was repeated in experiments with other dialects of Spanish (Argentinian and Cuban) employing electropalatography (Kochetov & Colantoni, 2011). Zsiga (1993) proposes that both gradient and categorical effects can be seen but that both rely on gestural representations.

Similar to the Spanish /n/, the Japanese mora nasal /ɴ/, which occurs in syllable-final position, takes its place of articulation from the following segment (Amanuma, Otsubo, & Mizutani, 1978; Nakajo, 1989; Vance, 2008). However, unlike Spanish, the mora nasal in utterance-final position is often transcribed as velar, uvular or even placeless. The present study examines the tongue shapes of a speaker of Tokyo Japanese using ultrasound imaging to investigate whether Japanese mora nasal /ɴ/ assimilation is gradient or categorical.

Figure1 compares speculative schematics of gestural scores of nasal assimilation from Honorof's Castilian Spanish study and Japanese data of this present study. The shaded sections show deleted gestures. As seen in the left panel of Figure1, the alveolar gesture for the nasal is deleted, and the lip gesture for the labial is extended in a sequence of /np/ in Castilian Spanish.

Figure 5: Gestural scores of Castilian Spanish and Japanese nasal assimilation. The shaded sections show deleted gestures.



If categorical assimilation occurs in different-articulator sequences, a velar or uvular gesture for the Japanese mora nasal could also be deleted before a labial. However, ultrasound imaging data from one speaker showed the velar gesture for /ɴ/ was not deleted, consistent with the central panel of Figure 1. In other words, all gestures remained and assimilation was not categorical in this case, even though perceptually, it was. The right panel shows that the velar gesture for /ɴ/ seemed to be deleted before an alveolar /n/, but this case could be a blending of the two tongue gestures, so that the presence of the blended gesture is harder to determine. Gestural blending was also observed in same-articulator sequences of /ɴg/. In sum, categorical assimilation was not observed in Japanese mora nasal /ɴ/ assimilation in any phonological environment in this study. The mora nasal may be variable across speakers (Yamane & Gick 2010), so further research is needed to determine whether all realizations of the mora nasal behave similarly.

References

Amanuma, Y., Otsubo, K., and Mizutani, O. (1978). *Nihongo onseigaku*. Tokyo: Kuroshio Shuppan. / Browman, C. P. and L. Goldstein (1989). Articulatory Gestures as Phonological Units. *Phonology* 6(2): 201-251. / Browman, C. P. and L. Goldstein (1990). Tiers in Articulatory Phonology, with Some Implications for Casual Speech. In John Kingston & Mary E. Beckman (eds.), *Papers in laboratory phonology I: Between the grammar and the physics of speech*, 341-376. Cambridge: Cambridge University Press. / Honorof, Douglas N. (1999). Articulatory gestures and Spanish nasal assimilation. New Haven, CT: Yale University Ph.D. dissertation. / Kochetov, A. and L. Colantoni (2011). Spanish nasal assimilation revisited: A cross-dialect electropalatographic study. *Laboratory Phonology* 2(2): 487-523. / Nakajo, O. (1989). *Nihongo no onin to akusento* (Dai 1-han. ed.). Tokyo: Keiso Shobo. / Vance, T. J. (2008). *The sounds of Japanese*. Cambridge, UK ; New York: Cambridge University Press. / Yamane, N. and Gick, B. (2010). Speaker-Specific Place of Articulation: Idiosyncratic Targets for Japanese Coda Nasal. *Canadian Acoustics, 38(3)*, 136-137. / Zsiga, E. C. (1993). *Features, gestures, and the temporal aspects of phonological organization.* Unpublished Ph.D. dissertation, Yale University.

Aspiration metathesis in Andalusian Spanish: The role of place of articulation

Hanna Ruch

URPP Language and Space, University of Zurich

hanna.ruch@uzh.ch

This study explores pre- and post-aspirated stops in Andalusian Spanish. The aim is to model how pre-aspiration changes into post-aspiration by analyzing durational parameters for stops in two age groups of two Andalusian Spanish varieties. A further aim is to better understand the coordination of glottal and oral events.

Syllable final /s/ is usually weakened to [h] in Andalusian Spanish (e.g. *pasta* 'paste' [ˈpahta]). For Western Andalusian Spanish, post-aspirated stops have been reported (e.g. [ˈpat^{h}a] Torreira 2012; Parrell 2012), and an apparent-time study showed evidence of a sound change in progress from pre- to post- aspiration for /st/-sequences (Ruch & Harrington, 2014). As to modelling the change from pre-aspiration to post-aspiration, Torreira (2012) suggests post-aspiration is brought about by a realignment of the oral closing gesture with the glottal opening gesture, which might be accompanied by compensatory lengthening of the oral closure. Parrell (2012) considers the possibility that post-aspiration is the result of a categorical shift from anti-phase to in-phase, i.e. from an asynchronous timing of the glottal opening with the oral closing gesture to a synchronous timing. While Parrell's study considers only /st/ sequences, Torreira's does include also velar and bilabial stops, but without testing explicitly the effect of place of articulation.

If the aspiration metathesis [hC] → [C^h] is the result of one single process, we expect the VOffT (voice offset time, i.e. pre-aspiration) shortening to be coupled with VOT (voice onset time, i.e. post-aspiration) lengthening across all three stop types. We further hypothesized that in Andalusian /s/ + voiceless stops sequences VOT would be longest for velar, intermediate for dental and shortest for bilabial stops; VOffT was expected to be longest for velars and shortest for bilabial stops. We analyzed 12 words with medial /sC/ (e.g. *espada, estado, escapa*), and 6 with medial /C/ (e.g. *separa, etapa, secaron*) produced by 48 speakers (18 words × 3 repetitions × 48 speakers = 2592 tokens). The 48 speakers comprise 24 native speakers of Eastern (EAS) and 24 of Western Andalusian Spanish (WAS), drawn from two age groups for each variety (see Ruch & Harrington, 2014). For each V1(s)CV2 sequence (e.g. /espa/ in /esˈpada/, /epa/ in /seˈpara/), boundaries were set manually for the onset of the preceding vowel (V1), closure onset and offset, and the offset of the subsequent vowel (V2). An acoustic procedure (Ruch & Harrington, 2014) was then used to set automatically the offset of voicing in V1 and the onset of voicing in V2, which were interpreted as the onset of VOffT and the offset of VOT.

Mixed models were applied to /sC/-words with either VOffT or VOT as the dependent variable, age, variety and place of articulation as fixed factors and word and speaker as random factors. Age showed a significant effect ($p < 0.01$) on VOffT, confirming previous findings of a sound change in progress. There was a significant effect of place of articulation ($p < 0.001$) and no interaction with age or variety, indicating that pre-aspiration is disappearing equally across stop types. VOffT was longest for velars and shortest for bilabials; the difference between dentals and bilabials was not significant. For VOT there was a significant three-way interaction between the fixed factors. Tukey-tests revealed that the difference between older and younger speakers was more marked in WAS ($p < 0.001$) than in EAS ($p < 0.05$). While EAS and older WAS speakers showed the expected VOT pattern (k > t > p), there was no significant difference between velar and dental stops among younger WAS speakers. The largest difference in mean VOT between younger and older speakers was found for /st/ (22.3 ms; /sk/ 11.4, /sp/ 10.8 ms). This indicates the metathesis from pre- to post-aspiration starts in medial /st/ and is only later generalized to /sp, sk/. Further durational analysis of the preceding vowel, offset and onset of voicing, and oral closure indicated that the oral gesture varies across variety, age group and stop type in its relative timing and in its duration. Total duration (i.e. VOffT + closure duration + VOT) was very similar across contexts and suggests that the glottal gesture remains relatively unchanged. /C/-words showed the expected VOT pattern (k > t > p) in all four speaker groups, and no effect of age or variety was found. /sC/-words differed from /C/-words in terms of VOffT, closure duration and total duration; for younger speakers in all stop contexts and for older WAS speakers in velar and dental stops also in terms of VOT.

Our apparent-time data are only partly compatible with the existing models of the Andalusian sound change from pre-aspiration to post-aspiration. The different patterns for pre- and post-aspiration and the interaction between speaker age and stop type indicate post-aspiration lengthening and pre-aspiration shortening are gradual processes that are to a certain degree independent from one another. The consistently long closure duration in /sC/ compared to /C/-words and the relatively short VOffT in older EAS speakers indicates pre-aspiration shortening and compensatory lengthening of the oral closure were previous to the emergence of post-aspiration. An aspiration metathesis can thus be the result of several gradual processes.

References

Parrell, B. 2012. The role of gestural phasing in Western Andalusian Spanish aspiration. *JPhon* 40, 37–45. / Torreira, F. 2012. Investigating the nature of aspirated stops in Western Andalusian Spanish. *JIPA* 42(1), 49–63. / Ruch, H. & Harrington, J. (2014). Synchronic and diachronic factors in the change from pre-aspiration to post-aspiration in Andalusian Spanish. *JPhon* 45, 12-25.

On /tl, dl/-misperception:
Language Specificity and Cross-linguistic Tendencies

Yuriko Matsumoto-Yokoe

Phonetics Laboratory, Doctoral Program in Linguistics, Sophia University, Tokyo

yuriko.matsumot@gmail.com

Typologically detested coronal-liquid clusters /tl/ and /dl/ have been reported to be misperceived as /kl, gl/ respectively by native English and French listeners (Hallé et al., 1998). According to their view, this phenomenon is induced by the fact that the problematic clusters represent the gaps in English/French inventories, and thus are prone to be phonotactically assimilated to the existing clusters.

Anonymous (submitted) presented a piece of counterevidence to the “phonotactic assimilation” view by conducting a similar perception experiments with Japanese listeners, who exhibited a robust tendency to avoid /tl, dl/ clusters. While it is considered impossible for the Japanese speakers to naturally acquire the knowledge to make a distinction between obstruent-liquid clusters, the result obtained for six such clusters (/pl, bl, tl, dl, kl, gl/) was highly asymmetrical. For example, the ratio of correct identification mounted to 72%, 66%, 63%, 58%, respectively, for /pl, bl, kl, gl/, whereas only 13% and 2% of /tl, dl/ clusters were correctly identified. Given this result, Anonymous (submitted) suggested that the phenomenon at hand should not be understood as a mere case of language-specific phonotactic assimilation; rather, it should be recognised as a universal tendency.

/tl, dl/-misperception, however, cannot be solely regarded as a universal phenomenon. Anonymous (submitted) reported that the vector of misperception differed according to the listeners’ linguistic background. Traditionally, as Hallé et al. first pointed out, /tl, dl/ are erroneously perceived by native English listeners as /kl, gl/, with the first segments’ place of articulation being transposed from coronal to velar. While it also held true for the English listeners in the current study, native Japanese listeners did not follow the suit. Instead they opted for misperceiving /tl, dl/ as /pl, bl/, respectively. The tendencies were robust in both language groups: as for /tl/, 61% of the English listeners labelled it /kl/, when 67% of the Japanese listeners answered /pl/. As for /dl/, 71% of the English listeners heard /gl/, whereas 51% of the Japanese listeners responded with /bl/.

Furthermore, it shall be worth reporting that yet another group of listeners – highly proficient bilinguals of Japanese and English – exhibited somewhat “neutral” results between the two monolingual groups. Of all 108 tokens of /tl/-stimuli, 38% of the bilingual responses indicated that they misperceived /tl/ as /pl/ (resembling the Japanese group, the “labial shift”). At the same time, another 34% was labelled /kl/ by the same bilingual participants (resembling the English group, the “velar shift”). Similarly, for /dl/-stimuli, 31% of the bilingual responses exhibited the Japanese-type identification (i.e., /bl/), while 45% followed the English-type (i.e., /gl/). None of the bilingual participants showed a complete bias towards either of the two languages. All bilingual participants showed more or less equitable tendencies, without consistently being “fully English” or “fully Japanese”.

It is apparent from these results that the direction of misperception is determined by the listeners’ linguistic background, and that it is of a language-specific nature. Acoustic analyses of the stimuli were conducted to see if certain acoustic cues are relevant to the problem at hand. Although none of the values related to the bursts offered plausible explanations, the second and the third formant transition patterns seemed to provide a clue for the Japanese listeners’ “labial shift”. To illustrate, /pl/ and /tl/ had flat/falling formant patterns. On the other hand, /kl/ showed rising formants. It can be hypothesised that the Japanese listeners use formant transitions as the primary cue for the place of articulation of a plosive before a lateral. By being exposed to a second language (in this case, English), bilingual speakers might have acquired a new standard (or a cue) for the detection of the place.

References

Anonymous. Perception of plosive-lateral clusters: A summary of experiments. Submitted. / Berent, I., Steriade, D., Lennertz, T., Vaknin, V. 2008. What we know about what we have never heard: Evidence from perceptual illusions. *Cognition*. 104(3), 591-630. / Hallé, P.A., Segui, J., Frauenfelder, U., Meunier, C. 1998. Processing of illegal consonant clusters: A case of perceptual assimilation? *J. Exp. Psychology*. 24(2), 592-608. / Hallé, P.A., Best, C.T., Bachrach, A. 2003. Perception of /dl/ and /tl/ clusters: a cross-linguistic perceptual study with French and Israeli listeners. *Proc. 15th ICPhS Barcelona,* 2893-2896. / Hallé, P.A., Best, C. 2007. Dental-to-velar perceptual assimilation: a cross-linguistic study of the perception of dental stop +/l/ clusters. *J. Acoust. Soc. Am*. 121(5), 2899-2914.

Vowel Harmony in French: Investigating formant trajectories.

Duniec, Agnieszka & Crouzet, Olivier

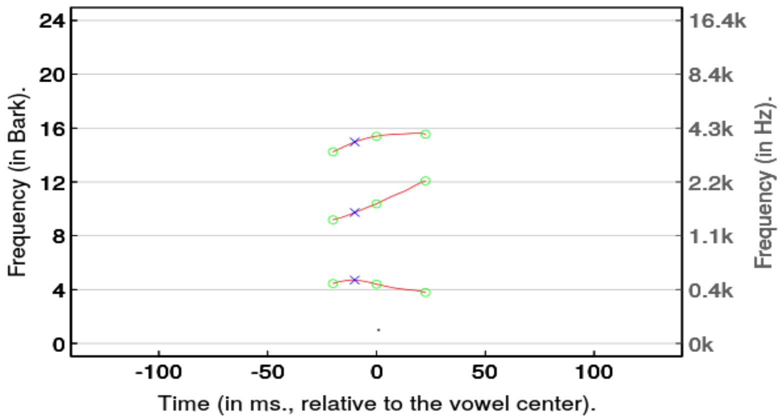
Laboratoire de Linguistique de Nantes – LLING / EA3827, Université de Nantes, Chemin de la Censive du Tertre, 44312 Nantes Cedex

agnieszka.duniec@etu.univ-nantes.fr|olivier.crouzet@univ-nantes.fr|

Vowel Harmony in French, as described by Grammont (1933, also see Fouché, 1956; Carton, 1974) has been investigated experimentally by Nguyen & Fagyal (2008). Acoustic analyses of the formant frequencies of vowels produced in [...]V1 CV2 [...] sequences showed that french mid-vowels ([e] ~ [ɛ], [o] ~ [ɔ], [ø] ~ [œ]) tend to exhibit a different position in the acoustic vowel space when followed by a non-low vowel compared to a mid-low or low vowel: when V2 is low (as in /eta/), V1tends to be lower than when it is followed by a non-low vowel (as in /ete/). This is apparent in acoustic measurements of both F1 and F2 formant frequencies and the statistical analyses that were performed by the authors are overall significant.

Though these data are in accordance with classical views on vowel harmony in French, Nguyen & Fagyal (2008)'s results were based on measurements taken from the acoustic midpoint of V1 vowels. It is usually acknowledged that this temporal midpoint may provide an adequate location for accurately measuring stable patterns of formant frequencies but there are alternative approaches to defining speech targets. One of them states that a better approach to locating the temporal moment where the speaker reaches the vowel static target is to identify the point where F1 reaches its maximum, which would correspond to the time when the vocal tract is maximally open (Lindblom & Sundberg, 1971). According to this proposal, if the time when F1 reaches a maximum precedes the temporal midpoint of the vowel, it would appear that the static vowel target has been reached before the temporal position at which the acoustic measurement has been taken (cf. Fig. 1 for an illustration). Of course, the temporal relationship between maximal F1 position and midpoint may vary. Nevertheless, it is crucial to investigate this issue as, in cases where F1 max would precede the vowel midpoint, evidence of a vowel harmony influence may appear to be the consequence of the formant movement trajectory between two unmodified targets and this may have influenced Nguyen & Fagyal (2008)'s data.

Figure 1: Schematic spectrogram of an actual pronounciation of V1 (3 formant tracks) in the word 'Monique' (/monik/) produced by Speaker 3. The temporal position when F1 reaches its maximum (identified by the cross) precedes the vowel midpoint (identified by the mid-circle). Measurements of F2 and F3 frequencies at these positions may obviously lead to alternative conclusions.



In order to address this issue, an experiment was designed to investigate vowel harmony phenomena in French at different temporal positions in the supposedly harmonised vowel in order to provide data on the time course of formant trajectories when V2 is either a low or a non-low vowel. Four French speakers took part in this experiment. Statistical analyses performed on the first three formants at 3 temporal positions within the first vowel partly confirm Nguyen & Fagyal (2008)'s conclusions. Issues relating to the time course of this phenomenon along the duration of the vowel are discussed according to theoretical approaches to static vs. dynamic vowel targets, calling for additional analyses and experiments.

References

Carton, F. (1974). Introduction à la phonétique du français. Paris: Bordas / Fouché, P. (1956). Traité de prononciation française. Paris, France: Klincksieck / Grammont, M. (1933). Traité de phonétique. Paris, France: Delagrave / Lindblom, B. E., & Sundberg, J. E. (1971). Acoustical Consequences of Lip, Tongue, Jaw and Larynx Movement. The Journal of the Acoustical Society of America, 50(4), 1166–1179 / Nguyen, N., & Fagyal, Z. (2008). Acoustic aspects of vowel harmony in French. Journal of Phonetics, 36(1), 1–27.

Japanese Nasal Place/Stricture Assimilation: Electropalatographic Evidence

Alexei Kochetov
University of Toronto
al.kochetov@utoronto.ca

Articulatory instrumental investigations of nasal place assimilation in a variety of languages revealed that the process can be realized categorically or gradiently – that is, showing a complete or partial constriction location identity between the nasal and the following consonant (Ellis & Hardcastle, 2002, among others). The studies also found that the process can apply differently before same-place consonants that differ in manner, showing either categorical or gradient assimilation in constriction degree – stricture (e.g. Farnetani & Busá, 1994). In an electropalatographic (EPG) study of Spanish (Kochetov & Colantoni, 2011), we found that speakers of two dialects showed similarly categorical patterns of nasal assimilation in *place*, but differed in their patterns of *stricture* assimilation. Specifically, Argentine nasals before /s/ and /ʃ/ were partly and variably deocclusivized, while Cuban nasals in the same contexts exhibited an invariant fricative-like central channel. One possible explanation for the dialect-specific patterns is the different phonetic realization of coda nasal – apical alveolar [n] in Argentine and velar [ŋ] in Cuban Spanish. The latter consonant, a result of historical and synchronic weakening, is possibly unspecified for constriction degree, and thus more prone to categorical assimilation.

If this hypothesis is correct, we would expect to observe similar patterns of categorical place-cum-stricture assimilation of Japanese moraic nasal /N/, which has been traditionally described as a velar or uvular of variable constriction degree (the utterance-final [ŋ̚], [ɴ], or [ɰ̃]; Vance, 1987), and analyzed phonologically as featurally underspecified (Kuroda, 1979). Descriptive accounts of Japanese do in fact mention that /N/ takes on the place of the following consonant, and, importantly, seems to lack a closure before fricatives (Vance, 1987). To our knowledge, the only articulatory investigation of Japanese nasal assimilation is Stevenson & Harrington (2002). Their EPG study confirmed the categorical status of Japanese place assimilation (found to be distinctly different from the largely gradient English place assimilation), but, being limited to the /N/ + /k/ sequence, did not explore the place and stricture relation.

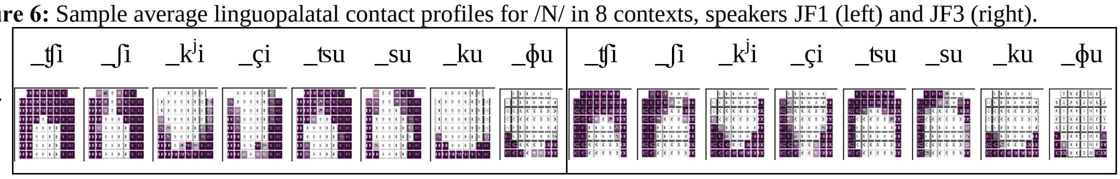
In this study we investigate Japanese nasal place/stricture assimilation in a variety of contexts. A WinEPG system with custom-made artificial palates was employed to collect simultaneous articulatory and acoustic data from 5 Japanese native speakers. To ensure that the assimilation patterns examined were fully productive, we used a language game (*Zuujago*: Itô et al., 1996) in which the speaker had to reverse morae within each word (e.g. /heN.te.ko na **kiː.ʧaN** no ma.neː.ʤaː/ 'Kichan's (name) weird boss' → /te.ko.heN na **ʧaN.ki**ː no ʤaː.ma.neː/). The resulting target words included /N/ followed by (palato-)alveolar affricates [ʧ] and [ʦ] (allophones of /t/), velar/palatal stops [k] and [k^j] (/k/), and palatal/bilabial fricatives [ç] and [ɸ] (/h/) before high vowels: [ʧaN+ʧi], [ʧaN+k^ji], [ʧaN+ʃi], [ʧaN+çi]; [ʧaN+ʦu], [ʧaN+ku], [ʧaN+su], and [ʧaN+ɸu]. 8 repetitions of each sentence were elicited in 2 separate sessions, producing 320 tokens in total. The analysis involved qualitative and quantitative measurements of assimilation taken at the midpoint of the acoustically determined nasal interval (cf. Stevenson & Harrington, 2002).

The results revealed consistent patterns of place assimilation: for all 5 speakers, the nasal was distinctly alveolar before [ʦ] and [s], palato-alveolar before [ʧ] and [ʃ], palatal before [k^j] and [ç], and velar before [k] and [ɸ] (see Figure 1). Importantly, context-specific constriction location differences were accompanied by constriction degree differences: the nasal had a stop-like closure before affricates and stops, and a fricative-like central channel before fricatives, also consistently for all speakers. While largely categorical, these place and stricture assimilation patterns showed some gradient effects, with constriction degree and location being somewhat less peripheral and less occluded than for the following consonants, particularly in the _i context. This observation lends some credit to Kuroda's (1979) impressionistic observation that the Japanese nasal is realized as a continuous transition from the preceding vowel to the following consonant. Overall, the results of the study contribute to the typology of nasal assimilation patterns, suggesting that the relation between place and stricture in nasals is more constrained and possibly conditioned by differences in gestural specification of coda nasals.

Figure 6: Sample average linguopalatal contact profiles for /N/ in 8 contexts, speakers JF1 (left) and JF3 (right).

(palato-)alveolar

palatal/velar



References

Ellis, L., & Hardcastle, W. (2002). Categorical and gradient properties of assimilation in alveolar to velar sequences: Evidence from EPG and EMA data. *JPhon, 30*. 373–396. / Farnetani, E., & Busà, M.G. (1994). Italian clusters in continuous speech. *Proc. 1994 ICSLP*, Vol. 1, 359–362. / Itô, J., Kitagawa, Y., & Mester, A. (1996). Prosodic faithfulness and correspondence: Evidence from a Japanese argot. *J. East Asian Ling., 5*, 217–294. / Kuroda, S. Y. 1979. *Generative grammatical studies in the Japanese language*. Garland. / Kochetov, A., & L. Colantoni (2011). Spanish nasal assimilation revisited: A cross-dialect electropalatographic study. *Laboratory Phonology*, 2, 487-523. / 2, 487-523. Stephenson, L.S., & Harrington, J. (2002). Assimilation of place of articulation: Evidence from English and Japanese. *Proc. 9th AICSST*, 592–597. / Vance, T.J. (1987). *An introduction to Japanese phonology*. U. of New York Press.

[DAY 3: Oral Session 9]

On establishing the existence of word stress

Carlos Gussenhoven

Radboud University Nijmegen

A language has word stress if a syllable-based culminative and obligatory prominence feature is part of the phonology of words (Hyman 2006). This definition excludes languages with a mora-based culminative and obligatory tone, like Kinga, languages with an obligatory phrase-based syllabic pitch accent, like French, and languages with non-obligatory syllable-based prominence, like Japanese.

A criterion not listed by Hyman is phonetic prominence. I will discuss data from a number of languages in which the relation between phonetic salience and stress is unexpected, confirming that phonetic salience measures do not define word stress.

In Ambonese Malay, a language without vowel quantity, minimal pairs like [barat] 'West' – [băraat] 'heavy' suggest that the language has word stress, but it is hard to make a case for its existence. The salient peaks would appear to be best analyzed as due to phrase-boundary melodies that remain floating. This position will be argued for on the basis of a peak alignment study. Second, while the status of word stress in varieties of Tamazight is ambiguous at best, in the Zuara variety penultimate stress is a regular feature of words, even those that have a voiceless obstruent in the rime of the penultimate syllable, like [a.ˈsq.qad] 'flail'. This position will be defended on the basis of acoustic measurements in questions and statements. Third, Standard Nigerian English has tonal structures which reflect the position of the word stress in British English. However, while other new varieties of English with tonal substrates, like Cantonese English, apparently lack word stress, Standard Nigerian English distinguishes words with initial and peninitial stress by means of duration as well as pitch. Its word prosodic structure will be argued for on the basis of acceptability judgements of sentential stimuli in which f0 has been manipulated.

An operational definition of word stress may be provided by the 'stress deafness' paradigm of Peperkamp & Dupoux and colleagues: if listeners perform poorly on reproducing the presentation order of series of stimuli that minimally differ in the position of phonetic prominence ([minú] – [mínu], …), the language doesn't have word stress. We ran a 'stress deafness' experiment to see whether Persian, which has generally been described as having contrastive word stress or accent, passes the 'stress' criterion, using languages that uncontroversially have lexical stress or accent and languages that don't as upper and lower baselines. The results suggest that the 'stress deafness' test discriminates between lexical and postlexical stress or accent, and that the reason that Persian listeners are 'stress deaf' is that their accent distinctions arise postlexically.

[DAY 3: Oral Session 10]

Pointed and plateau-shaped pitch accents in North Frisian dialects

Oliver Niebuhr[1] & Jarich Hoekstra[2]

[1] Dept. of General Linguistics, [2] Frisian Dept., ISFAS, Kiel University, Germany

niebuhr@isfas.uni-kiel.de; j.hoekstra@isfas.uni-kiel.de

Our paper summarizes the results of a pilot analysis exploring the intonational phonology of North Frisian. North Frisian is one of three main varieties of Frisian. Despite its diversity, Frisian may be considered an endangered language with overall less than 500,000 (mostly elderly) speakers who live along the coastline of the North Sea from the Netherlands via Germany to Southern Denmark. We focus here on two closely related North Frisian dialects, Fering and Öömrang, which are spoken by only a few hundred speakers on small islands off the German coastline. Analyses of the North Frisian intonation patterns are so far restricted to auditory descriptions that date back to the early 20th century (Tedsen 1906). So, modern analyses are urgently needed not least because North Frisian is on the retreat since the 17th century in terms of both territory and number of speakers.

Our pilot analysis started with an auditory observation. While the majority of pitch accents in Fering and Öömrang sound on the whole "ordinary" rising/falling to Standard German ears, some pitch accents are strikingly different in that they create the impression of being "halted" or "disfluent" due to a flattened peak maximum. That is, there seems to be an intonational contrast between pitch accent peaks with pointed and plateau-shaped maxima in North Frisian. Our primary aim was to support this auditory impression by acoustic-prosodic measurements, in this way also providing basic alignment details for the two types of pitch accents. Moreover, our secondary aim was to find out how the "halted/disfluent" sounding accents differ functionally from the "ordinary" sounding ones by analysing the semantic-pragmatic contexts in which they occurred.

The analysed speech data come from both read and spontaneous speech corpora recorded by the second author for fieldwork and language-documentation purposes. We measured established prosodic parameters of 120 pitch accents, half of which belonged to the "halted/disfluent" sounding type. The other half was a control sample with "ordinary" sounding H* pitch accents from the same speakers. The measurements included (1) vowel duration, (2) alignment of rise onset/offset, fall onset/offset, and F0 maximum relative to the accented syllable and vowel, (3) duration of peak maximum (in terms of 1 semitone to both sides of the maximum F0 value), as well as (4) the ranges of the rising and falling F0 movements. Significant differences were tested for in a one-way MANOVA.

Next to a number of interesting details, our prosodic analysis yielded two main results. First, as regards F0 shape and scaling, we found clear and significant evidence for a distinction between pointed and plateau-shaped pitch accents. The "ordinary" sounding pitch accents had pointed peak maxima of about 30-50 ms, whereas the "halted/disfluent" sounding pitch accents had plateau-shaped peak maxima with durations of at least 70 and up to 250 ms. The clear shape difference between the two types of pitch accents was not linked with significant differences in the F0 ranges of the rising or falling movements. Second, as regards alignment, significant correlations provide evidence for a close coordination between many F0 landmarks and segment boundaries, such as rise onset and accented-syllable onset, or peak maximum/fall onset and accented-vowel offset. Most remarkably, however, the plateau-shaped pitch accents were often timed such that they precisely spanned the entire accented vowel. That is, the plateau started at the vowel onset and ended at the vowel offset, independently of vowel quantity. The corresponding correlation is illustrated in Figure 1a. An example is given in Figure 1b.

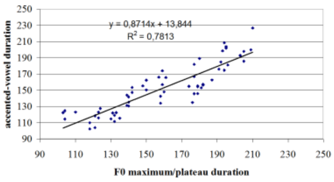
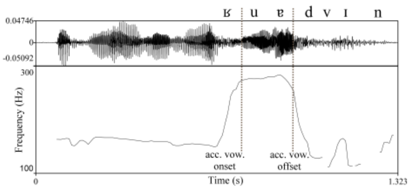


Figure 1a: Correlation of acc. vow. dur. and F0 max dur. (in ms), n=60 r=0.8, p<0.001

Figure 1b: Example of plateau-shaped F0 peak on [ʁuɐdvin] (red wine) in Öömrang



The functional analysis suggests that the plateau-shaped type of pitch accent occurs under emphatic conditions, including those of narrow focus. This functional role, in combination with the fact that there were no F0 range differences between the two types of pitch accents, lead us to the assumption that North Frisian speakers flatten rather than lift the F0 maximum in order to make the pitch accent sound higher and more prominent (cf. Knight, 2008). This assumed substitution strategy, which has to our knowledge never been noted before, is supported by further speaker-specific analyses. They indicate that many speakers applied both strategies, peak lifting and peak flattening, to different degrees hence creating a substitution continuum with 'lifters' and 'flatteners' at its extremes. According to recent findings, a similar pattern seems to exist in East Frisian Low Saxon (Gackstatter & Niebuhr, 2012), which is an endangered variety of Low German with regional and historical relations to Frisian. Besides their value for intonational typology, our findings can be discussed with respect to phonological modeling (tonal spreading/anchoring, individual trade-offs) and the theory of biological codes (Gussenhoven 2004).

References

Gackstatter, M., Niebuhr, O. 2012. Eine kontrastive phonetische Analyse Niederdeutscher Langvokale. *Linguistik Online 53*, 23-54 / Gussenhoven, C. 2002. Intonation and interpretation: phonetics and phonology. *Proc. 1st Speech Prosody, Aix-en-Provence, France, 47-57* / Knight, R.-A. 2008. The shape of nuclear falls and their effect on the perception of pitch and prominence: peaks vs. plateaux. *Language & Speech 51, 223-244* / Tedsen, J. 1906. Der Lautstand der föhringischen

Quantity Contrast in Lule Saami: A Three-Way System

Nora Fangel-Gustavson[1], & Bruce Morén-Duolljá[2]
[1] Laboratoire de Phonétique et Phonologie (CNRS/Sorbonne Nouvelle), [2] University of Tromsø
nora.fangel-gustavson@univ-paris3.fr, bruce.moren@uit.no

Lule Saami, a severely endangered indigenous language of Northern Scandinavia, presents a three-way consonantal quantity system that allows for both lexical and morphological contrast. Word-medial consonants can occur with three different degrees of length; quantities (Q) 1, 2 and 3, or singletons, geminates and supergeminates. We carried out a production study in order to investigate how this contrast is acoustically implemented. Specifically, we wanted to determine whether phonological quantity is associated with other cues other than the temporal ones, as suggested by earlier evidence for e.g. Estonian and Inari Saami (Lehiste 1997, Ehala 2003, Spahr 2012, Bye et al. 2010). Implications of the results for the general issue of geminate representation are discussed, with particular attention to the way moraic theory may account for the ternary system of Saami quantity.

Eight native speakers of Lule Saami were recorded in the field in Tysfjord (Norway). The corpus consisted of six verbs in three different forms (i.e. six minimal triples), allowing to measure alternations across the three quantities (Q1, Q2, Q3) in a **(C)V1CV2** context for consonants [l], [m], [n], [r], [s] and [v]. Data were segmented, annotated and analyzed using Praat (Boersma & Weenick 2012). Durations of each segment were measured and the non-temporal values extracted from the post word-initial onset material of each token were fundamental frequency, F1, F2, F3 (all in Hz) and intensity (dB).

Results show that the triple contrast is phonetically implemented through robust durational differences among the Q1, Q2 and Q3 consonants tested. The mean duration of a singleton (Q1) was 90 ms, a geminate (Q2) 215 ms and a supergeminate (Q3) 323 ms. There was a slight readjustment effect on adjacent vowels. The largest effect was on the vowel following the target consonant (V2). The average duration of the total target VCV sequence was systematically longer from Q1 to Q2 and from Q2 to Q3. As far as non-temporal parameters are concerned, speakers produced a high-low intonational pattern within the disyllabic sequence, with an f0 peak at the left-edge of the target consonant in all three quantities, with somewhat higher peak in Q2 and Q3 than in Q1. Quantity had a significant effect on adjacent vowel intensity (V1: p = .01, V2: p = .02) but not consonant intensity (p = 0.18). V1 intensity increased with quantity whereas V2 intensity decreased. Acoustic vowel formants were not affected.

Several works argue that phonetic duration is organized around the mora (Port et al. 1987, Hubbard 1995, Broselow et al. 1997, Ham 1998). Moraic theory (Hyman 1985, McCarthy & Prince 1986, Hayes 1989) represents geminates and supergeminates as melodic units linked to one and two morae, respectively. Supergeminates are problematic for the principle of Maximal Binarity (McCarthy & Prince 1986), which claims an upper limit of two morae per syllable. However, trimoraic syllables have been claimed to exist in several languages, including Hungarian, Estonian, Hindi and Persian (Ham 1998, Hayes 1989). A moraic analysis of Lule Saami supergeminates that adds a mora stepwise from Q1 → Q2 → Q3 fits nicely with the temporal differences observed in our study. Assuming this analysis entails linking the Q3 consonant to two morae, yielding a trimoraic first syllable. We will show, however, that this analysis fails to account for other, non-durational alternations present in Lule Saami, including changes in laryngeal features (for voicing and aspiration), suggesting that a full-fledged analysis accounting for both quantitative and qualitative alternations has yet to be found.

References

Bye, P., Sagulin, E., & Toivonen, I. (2010). "Phonetic duration, phonological quantity and prosodic structure in Inari Saami". *Phonetica*, 66(4), 199-221. / Broselow, E., S.-I. Chen & M. Huffman. (1997) Syllable Weight: Convergence of Phonology and Phonetics. *Phonology* 14, 47-83. / Ehala, M. (2003). Estonian quantity: Implications for moraic theory. *Generative Approaches to Finnic and Saami Linguistics. Stanford: CSLI, 5*, 79. / Ham, W. (1998). *Phonetic and Phonological Aspects of Geminate Timing*. Ph.D. dissertation, Cornell University. / Hayes, B. (1989). Compensatory lengthening in Moraic Phonology. *Linguistic Inquiry* (The Massachusetts Institute of Technology) 20 (2): 253–306. / Hubbard, K. (1995). Toward a theory of phonological and phonetic timing: Evidence from Bantu. *Phonological and Phonetic Evidence: Papers in Laboratory Phonology IV*. Cambridge: CUP, 168-187. / Hyman, L. M. (1985). *A theory of phonological weight*. Dordrecht: Foris. / Lehiste, I. (1997). Search for phonetic correlates in Estonian Prosody. In I. Lehiste, & J. Ross (Eds.), *Estonian prosody: Papers from a symposium* (pp. 11–33). Tallinn: Institute of Estonian Language. / McCarthy, J. and A. Prince. (1986). *Prosodic morphology*. Manuscript / Port, R. F., Dalby, J., & O'Dell, M. (1987). Evidence for mora timing in Japanese. *The Journal of the Acoustical Society of America, 81*, 1574. / Spahr, C. (2012). Rethinking the morphophonology of Estonian quantity. Nordic Prosody XI.

Prominence, Phrasing, and Information Structure in Mawng (Australian)

Janet Fletcher, Ruth Singer, Debbie Loakes
University of Melbourne
janetf@unimelb.edu.au;rsinger@unimelb.edu.au;dloakes@unimelb.edu.au

Relatively few studies have examined the prosody and information structure interface in Australian indigenous languages with some notable exceptions (e.g. Anonymous 2006, Hellmuth et al. 2007). In Mawng, the language under investigation in this study, corrective focus on objects can be expressed grammatically in at least two ways. The object must be overt (obligatory), and syntactic fronting may also be observed (less common). Syntactic fronting has also been noted in a range of other Australian languages (e.g. Warlpiri, Arrernte), as well as in many so called "free word order" languages in the world (e.g. see summary in Féry 2013). It is also well known that a range of intonational devices are used in the grammar of pragmatic focus marking in languages. These might include combinations of the following: manipulations of phrase-level pitch range, intonational and prosodic phrasing, and intonational prominence, including the use of different types of pitch accents for contrastive emphasis. Languages can also de-accent material (reducing the number of pitch-accents in a phrase) and/or de-phrase non-focal material (reducing the number of intonational constituents) to promote a particular kind of discourse interpretation. This also fits with a general observation that languages with diverse prosodic characteristics like Hindi, Mandarin, or Chichewâ also tend to show post-focal pitch range compression, with prosodic phrasing or lexical tonal contrasts (in the case of Mandarin and Chichewâ) left intact. Although Mawng has been analysed previously as a stress language (Anonymous 2006), it remains to be seen how prosodic phrasing and pitch range variation contribute to focus marking in this language or how post-focal material is realized prosodically across different discourse types.

Two separate corpora of Mawng, an endangered language spoken on Goulburn Island off the coast of Northern Australia, were analysed in this study. The first corpus is a controlled read speech experiment that was designed to explore prosodic realisation of contrastive or "corrective" focus on fronted nominal objects in Mawng mini-dialogues. The second corpus consists of a series of questions and answer pairs elicited using a selection of QUIS ("Question and Information Structure") materials developed at Potsdam University that were designed to elicit different focal contexts (e.g. see Skopoteas et al, 2006). Three speakers were recorded in the first experiment, and three speakers participated in the second experiment. An example of a dialogue from Experiment 1 is shown below:

Speaker 1:	*Arriwarnangajpun Walmuri*.	"We call it Pufferfish."
Speaker 2:	*Marrik arriwarnangajpu Walmuri.*	"We don't call it Pufferfish."
	Alngkat *arriwarnangajpun.*	"We call it **Stonefish**."

Word order was deliberately manipulated in this experiment to elicit corrective focus on the initial noun object in the final utterance of the dialogue. In the second corpus speakers responded to picture prompts and were not required to read a set of fixed materials. Approximately 600 mini-dialogues were analysed for the first experiment and 170 question/answer pairs from the second corpus.

Results of Experiment 1 showed that all three speakers produced fronted focal elements with a major pitch movement associated with the focused word, which was analysed as a H* or L+H* pitch accent often realised in the highest part of the speaker's range. There was no evidence that a phonologically different kind of focal pitch accent from a "regular" pitch accent was used by our three speakers to indicate corrective focus in these data, unlike in Bengali, for example, in agreement with earlier analyses of Mawng (e.g. Hellmuth et al. 2007). Post-focal material was almost always produced in a relatively compressed pitch range but with no evidence of post-focal de-accentuation. Rather, fronted elements were often realised in their own prosodic phrase. In other words there was a clear downwards re-setting of pitch range on the subsequent prosodic phrase regardless of whether there was the presence of a pause or major intonational boundary separating the fronted focal element and the rest of the utterance. In the second corpus, more variation was apparent given the less constrained nature of the speech task, but overall similar patterns of pitch range manipulation and phrasing were observed in corrective focus contexts. Focused objects preceding or following a verb were always accented and often realised in a separate prosodic phrase, mostly with boosted pitch range. However variable pre-focal pitch range compression was observed, whereas post-focal material was almost always produced in a significantly lower pitch register concurring with patterns noted in a range of other typologically diverse languages.

References

Féry, C. 2013. Focus as prosodic alignment. *Nat lang Linguist Theory* 31 (3), 683-734 / Hellmuth, S., Kügler, F. & Anonymous. 2007. Quantitative investigation of intonation in an endangered language. *Proc. of conference on language documentation and linguistic theory*, In Austin, P., Bond, O. & Nathan, D. (eds.) London: SOAS. 123-132./ Anonymous 2006. Information structure in Mawng: intonation and focus. *Selected papers of the 2005 conference of the Australian Linguistics Society*, Allan, K. (ed). http://www.als.asn.au/. Skopeteas, S., Fiedler, I. Hellmuth, S., Schwarz, A., Stoel, R., Fanselow, G., Fery, C., & Krifka, M.(eds.) (2006).*Questionaire on information structure (QUIS)*. Potsdam: University of Potsdam. http://www.sfb632.uni-potsdam.de/quis.html

No VOT perception without native VOT experience

Rikke Bundgaard-Nielsen & Brett Baker
La Trobe University/University of Western Sydney, University of Melbourne
rikkelou@gmail.com, bjbaker@unimelb.edu.au

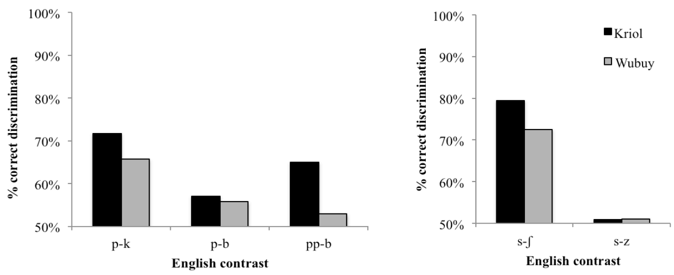
Background. 50 years of speech perception research provides a rich literature on cross- and second language (L2) perception of stop consonant contrasts such as /p b/, /t d/, and /k g/ which differ systematically in the relative timing of oral stop release and the onset of vocal fold vibration (voice onset time: VOT). This research has focused primarily on two observations: 1) that nonnative listeners automatically use their native VOT contrast boundary in perceiving phones in an unfamiliar language, and 2) that even highly proficient L2 language users often perceive L2 VOT-based contrasts in a way that is consistent with their L1, even after decades of L2 acquisition. No work has, hitherto, examined VOT-based contrast discrimination by L1 speakers of languages which lack VOT-based contrasts altogether. In the following, we present two studies of speakers in such a scenario.

Method. We tested the discrimination of English stops and fricatives by two groups of listeners in the remote Aboriginal settlement of Numbulwar in Arnhem Land (NT, Australia), who differ systematically in their native VOT experience: 8 speakers of the Indigenous language Wubuy, which is one of the atypical languages of the world without a VOT-based stop contrast and no fricatives, and 9 speakers of Roper Kriol (an English-lexified creole), which does have a voicing distinction in stops but not in fricatives, and in which the contrast for intervocalic stops is also realised through systematic differences in constriction duration (CD) (Baker et al., in press). All speakers had some competence in the 'other' community language (i.e., Roper Kriol for Wubuy speakers, and Wubuy for Kriol speakers), and all spoke some English. Study 1 tested discrimination of English intervocalic stops /p k/, /p b/, and a manipulated contrast of /p b/ where the CD of /p/ had been artificially lengthened to reflect the ratios in Roper Kriol (we denote this /pp/). Study 2 tested discrimination of syllable-initial English fricatives /s ʃ/ and /s z/. A control group of 12 English speakers had a mean discrimination accuracy ranging between 94-97% for all contrasts.

Predictions. On the basis of PAM/PAM-L2 assimilation patterns (Best, 1995; Best & Tyler, 2007), we predict that: **P1:** Wubuy speakers perceive /p, pp, b/ all as instances of native /p/ and fail to discriminate them. Kriol speakers perceive /p, pp/ as instances of Kriol /p/, and /b/ as Kriol /b/ and discriminate them though /p b/ will be discriminated less successfully than /pp b/ due to the lack of native Kriol-like CD differentiation. **P2:** Wubuy speakers cannot discriminate /s z/ as both will be uncategorisable phones and they have no native experience with voicing distinctions. Kriol speakers perceive /s z/ as instances of Kriol /s/ and cannot discriminate them. **P3:** discrimination of /s ʃ/ may be fair to good as both groups have extensive experience with multiple place of articulation contrasts in the alveolar region (Wubuy has a four-way coronal contrast in stops, nasals, and laterals: Ladefoged 2001. Kriol has an /s ʃ/ contrast.). **P4:** Control contrast /p k/ will be well-discriminated by both groups, as it corresponds to native categories.

Results and Discussion. Our results (Figures 1a and 1b) show that Wubuy speakers cannot discriminate VOT-based stop contrasts, even when the /p b/ contrast is acoustically enhanced by introducing a systematic difference in CD. Kriol speakers also find VOT-based stop contrasts very difficult to discriminate when they are not coupled with systematic differences in CD, as is the case in their L1. Both speaker groups fail to discriminate the VOT-based fricative contrast /s z/. Interestingly, both Wubuy and Kriol speakers show sensitivity, and reasonable discrimination performance, to place-of-articulation differences for English fricatives /s ʃ/. Our findings demonstrate that even extensive L2 learning and continued L2 use does not automatically allow learners to acquire an L2 VOT distinction when there is none in the L1.

Figure 1a and 1b. Mean Kriol and Wubuy discrimination accuracy in Study 1 (stops) and study 2 (fricatives).



References

Baker, B., Bundgaard-Nielsen, R., & Graetzer, S. (in press). The obstruent inventory of Roper Kriol. *Australian Journal of Linguistics*. / Best, C. T. (1995). A direct-realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience* (pp. 171-204). Timonium, MD: York Press. / Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In J. Munro & O.-S. Bohn (Eds.), *Second language speech learning* (pp. 13-34). Amsterdam: John Benjamins. / Ladefoged, P. (2001). *Vowels and Consonants*. Oxford: Blackwell.

[DAY 3: Oral Session 11]

Articulatory Correlates of Phonological Relationships

Kathleen Currie Hall, Hanna Smith, Kevin McMullin, Noriko Yamane, Blake Allen, Joash Gambarage
University of British Columbia
Corresponding author: kathleen.hall@ubc.ca

This paper has the following aims: 1) To provide evidence that, despite the common belief that perceived differences between contrastive and allophonic sounds are purely cognitive, there are in fact articulatory differences between them as well; 2) To provide evidence that such differences are sensitive to gradient degrees of contrast; and 3) To introduce the LabPhon community to optical flow analysis – a relatively novel technique for analyzing ultrasound data that allows for the collection of large amounts of data and for direct comparison across different speakers.

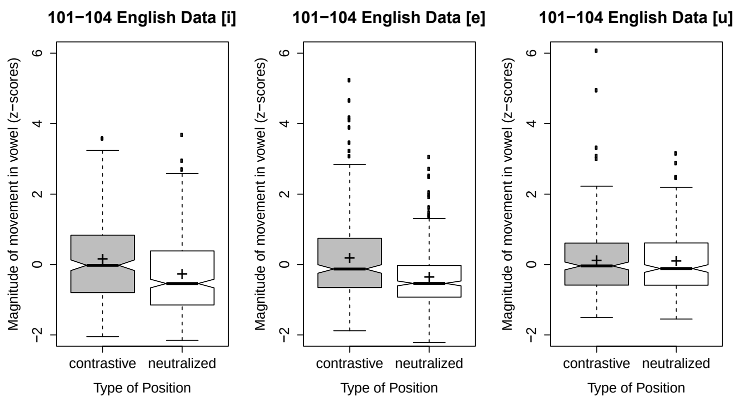
BACKGROUND: It is well established that sounds that are contrastive are often perceived as being more distinct than sounds that are not contrastive in a language (e.g., Ohala 1982, Whalen et al. 1997, Hume & Johnson 2003, Boomershine et al. 2008, Babel & Johnson 2010). The received wisdom is that such effects reflect the perception and processing of language by the minds of listeners, but there is no effect on language production (e.g., Kazanina et al. 2006, Boomershine et al. 2008). A few studies suggest that differences between contrastive and non-contrastive sounds can be enhanced in special conditions (e.g., in competition with minimal pairs (Baese-Berk & Goldrick 2009) or in child-directed speech (Cristia & Seidl to appear)). In this paper, we present novel results that suggest that there are in fact articulatory differences in regular adult speech that are dependent on the degree of contrast of various sounds. Furthermore, the results support the hypothesis that there are subtle differences in degrees of contrastivity, rather than simply a binary split between contrastive and non-contrastive (cf. Ladd 2006, Scobbie & Stuart-Smith 2008, Hall 2013).

METHODOLOGY: Ultrasound video imaging was used to collect extensive motion data for productions of tense and lax vowels in English. Tense vowels tend to contrast with lax vowels in closed syllables (e.g., *bead* vs. *bid*), but the contrast is neutralized in word-final open syllables (e.g., *bee* but *[bɪ]). Thus, if degree of contrastiveness affects production, one would expect to see larger differences in word-final closed syllables than in word-final open syllables.

Optical flow analysis (e.g., Horn & Schunck 1981, Barbosa et al. 2008) was applied to the video imaging. Optical flow has been used for decades in video research and allows one to explore general differences in velocity and magnitude of movement. Applying it to ultrasound video allows more data to be collected than standard tongue-contour analysis, both because it uses information from an entire utterance and because it is easy to extract flow data from many speakers. Furthermore, the resulting data can be directly compared across speakers using a standard normalization (e.g. a z-score).

RESULTS: Results from four speakers have so far been analyzed. For [i] and [e], there are clear and significant differences as expected, with vowels showing greater movement from frame to frame in contrastive positions than in non-contrastive ones. This effect is consistently not found with [u], which can be attributed to the fact that [u] and [ʊ] are much *less* contrastive in closed syllables than the front vowel pairs. Using the measure of predictability of distribution proposed in Hall (2009), where complete predictability (allophony) is measured as having an entropy of 0 and complete unpredictability (contrast) as an entropy of 1, [i]/[ɪ] and [e]/[ɛ] in word-final closed syllables have entropies of 0.95 and 0.995, respectively, while [u]/[ʊ] has a much lower entropy of 0.7. Thus, not only does it seem that speakers may be making articulatory differences based on phonological relations, but they also seem to be sensitive to gradient distinctions in those relations. These results lend credence both to the psychological reality of the notion of contrast as a fundamental construct in phonology as well as to the possibility that contrast is not a binary measure.

Figure 1: Averaged results from four English speakers



References
Available at: http://anonymousreferences.wordpress.com/2014/01/31/abstract-articulatory-correlates-of-phonological-relationships/

Invariant coupling relations at the core of variable speech trajectories

Leonardo Lancia
Max Planck Institute for Evolutionary Anthropology
leonardo_lancia@eva.mpg.de

In the last 20-30 years several researchers adopted the view that a speech gesture corresponds to a dynamical system governing the behavior of an abstract variable. This may have an articulatory nature, as in articulatory phonology, but in principle it can also correspond to an acoustic parameter, or to a combination of acoustic and articulatory parameters. The dynamical system controlling the abstract variable is implemented by setting coupling relations which tie the behaviors of the dynamical systems governing the various physical variables involved in the production of the gesture. Therefore we can in principle determine which are the abstract variables controlled speech production by studying the coupling relations underlying the dependencies between the trajectories of the physical parameters involved the production of speech gestures. Available methods are either linear, like Granger causality (clearly inadequate to detect nonlinear relations) or are derived from information theory (and necessitate large amounts of data points per trajectory). Here we propose a new approach based on joint recurrence analysis (JRA) to quantify the strength of coupling, and on the mean conditional probabilities of recurrence (MCR) to determine the direction of coupling (i.e. which system influences the other). These methods are based on the idea that, if the presence of perturbations in one time series determines the presence of perturbations in another time series, the system governing the first has an influence on the system governing the second (Zou et al, 2011). Thanks to the role played by the variability of the observed trajectories, this approach also has a theoretical advantage over competing approaches. Indeed, the coupling relations between the dynamical systems involved in the production of a given gesture permit to reduce the variability of the controlled abstract variable by distributing it on the involved physical variables (Shöner, 1995). We adapted these methods by following Lancia et al (in press) in such a way that their behavior do not depend on variations of the rates of change of the trajectories studied. This is a crucial feature of the proposed method because the rate of change of speech signals do not depend solely on the dynamics of the segmental speech gestures, but also on other linguistic (e.g. prosodic) and extra linguistic factors. Although this new approach could be adopted to study various kinds of speech-dependent variables (eg, acoustic or aerodynamic parameters), here we analyze articulatory data collected by means of electromagnetic articulography (EMA). In these data speech rate is highly variable and therefore it is possible to test the claim that the method proposed does not depend on variations of the rate of change of the analyzed signals. Indeed the data analyzed were produced by 4 German speakers and 5 French speakers who were asked to repeatedly utter CVCV utterances containing a labial and a coronal consonant while increasing their speech rate. The signals analyzed here represent the vertical motion of metallic pellets glued on the tip of the tongue (ttip), on the lower lip (llip) and below the lower incisors (to track jaw movements). In precedent works it was observed that, while at slow speech rate the jaw is raised each time the llip or the ttip are raised; at fast speech rate only one upward movement of the jaw is observed per utterance and this was mostly synchronous with the ttip, as if the llip was decoupled from the jaw. By applying our version of JRA to pairs of articulatory trajectories recorded during the repetition of each consonant, we obtained a positive definite measure of the strength of coupling (coupling is detected if the measure is higher than zero). By means of the modified MCR we obtained a measure of the direction of coupling centered around zero (no articulator drives the other if the measure is not different from 0). We considered these measures as dependent variables in two mixed effects models (one for the strength and one for the direction of coupling). The predictors were the pair of articulators compared, the place of articulation of the consonant, the order of the consonants in the utterance to repeat, the duration of the utterances and all possible interactions. A maximal random effect structure was adopted. Results indicate that despite qualitative differences in the shapes of the trajectories due to speech rate, the coupling relations are invariant to speech rate while they depend on the place of articulation. During the production of the coronal consonants, the movement of the jaw and that of the llip depend on the movement of the ttip; while during the production of the labial consonants, the movement of the jaw and that of the ttip depend on the movement of the llip. These results, together with analyses of synthetic trajectories in which coupling relations were systematically manipulated, prove the reliability of our approach and provide evidence that qualitative differences in articulatory data can emerge from the interactions between physical contingencies and the invariant coupling relations defining speech gestures. The analytical power of this approach can be further enhanced in the study of natural corpora by constructing surrogate time series with the same characteristics of the relevant time series but lacking the expected coupling relations.

References

Lancia, L., Fuchs, S. & Tiede, M. (in press). Applications of concepts from cross-Recurrence Analysis in speech production: An overview and a comparison to other nonlinear methods, accepted in JSLHR. / Schöner, G. (1995). Recent developments and problems in human movement science and their conceptual implications. Ecological Psychology, 7(4), 291-314. / Zou, Y., Romano, M. C., Thiel, M., Marwan, N., & Kurths, J. (2011). Inferring indirect coupling by means of recurrences. International Journal of Bifurcation and Chaos, 21(04), 1099-1111.

Acoustic variability aids the interpretation of phonetic detail in cross-language speech production

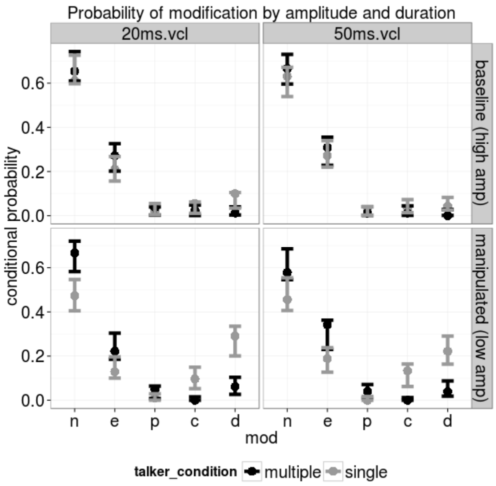
Lisa Davidson[a], Sean Martin[a], and Colin Wilson[b]
[a]New York University, [b]Johns Hopkins University
lisa.davidson@nyu.edu, sean.martin@nyu.edu, colin.wilson@jhu.edu

Speakers attend to fine-grained phonetic detail when attempting to produce words containing unfamiliar phonological structures (Anonymous, under revision). Instead of simply mimicking this detail, speakers appear to interpret the relevant acoustic cues in a way that is consistent with their native sound system. For example, English speakers may interpret bursts in stop-obstruent onsets (e.g. /bdafa/) as reduced vowels, because short high intensity transitions between two stop closures resemble possible phonetic implementations of English schwa (e.g., *p[ə]tato*, Davidson 2006). While such attention to fine-grained detail may be useful for learning new sound categories, it can also make non-native listeners overly sensitive to phonetic properties that are not contrastive in the target language. The aim of the current study is to examine whether providing speakers with phonetic variation, in the form of productions of a word by multiple talkers, aids them in establishing more stable representations of non-native sound sequences (cf. similar findings for phonemes by Lively et al. 1994, Bradlow et al. 1997). We hypothesize that hearing renditions of the same sequence by multiple talkers will aid non-native listeners in separating phonetic properties that are phonologically contrastive from those that are due to idiosyncratic phonetic implementation.

In this experiment, English speakers (N=48) repeated nonword stimuli, recorded by native Russian speakers, beginning with four types of C1C2 clusters: stop-stop (e.g., /bdafa/), stop-nasal (/tmado/), fricative-stop (/zgade/) and fricative-nasal (/zmaku/). To investigate how stimulus-specific detail affects English speakers' modifications of these clusters, three acoustic cues were systematically manipulated: (1) stop bursts were either 20ms or 50ms in duration; (2) the amplitude of the stop burst was raised or lowered relative to natural values; and (3) voicing of the initial consonant either began simultaneously with frication/stop closure or preceded the constriction (pre-obstruent voicing, POV). In the *single talker* condition, each of the manipulated stimuli were presented alone, while in the *multiple talker* condition the manipulated stimuli were preceded by baseline recordings of the same word by two different talkers.

Results from the single-talker condition show that English participants are extremely sensitive to the acoustic manipulations: there was significantly more epenthesis for 50ms stimuli; the amplitude manipulation led to more epenthesis for higher-amplitude bursts, and more C1 deletion and C1 change for lower-amplitude bursts; and there was significantly more prothesis when POV was present. However, the effects of these manipulations were significantly attenuated in the multiple-talker condition. The presence of POV still resulted in significantly higher rates of prothesis, but the effects of the duration and amplitude manipulations disappeared (note that modifications of epenthesis are common regardless of manipulation and condition). Fig. 1 illustrates the difference between conditions with the production results for voiceless stop-stop clusters. Importantly, such differences cannot be due to lack of attention to the manipulated stimulus in the multiple talker condition, as the burst durations of English speakers were significantly longer for the 50ms stimuli in both conditions.

Figure 1: Effect of manipulations on voiceless stop-stop sequences. n=no modification, e=epenthesis, p=prothesis, c=C1 change, d=deletion



Presenting speakers with multiple acoustically different instances of a non-native sequence leads to attenuation of their sensitivity to phonetic detail in production. That the effects of the manipulations are not completely eradicated suggests that speakers are *blending* all of the information they perceive, perhaps with a different weight assigned to each stimulus and with a preference for preserving the phonemes that are perceived. Regardless of the precise implementation of blending, the current results establish that providing non-native speakers with stimuli from multiple talkers prevents them from over-interpreting subphonemic phonetic detail in the target language. Talker variation helps the non-native speaker determine which phonetic properties are contrastive for new sound structures.

References

Anonymous. under revision. Language-specific phonetic detail constrains cross-language speech production. / Bradlow, A., et al. 1997. Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *J. Acoust. Soc. Am*. 101, 2299-2310. / Davidson, L. (2006). Schwa elision in fast speech: segmental deletion or gestural overlap? *Phonetica* 63:2-3, 79-112. / Lively, S., et al. 1994. Training Japanese listeners to identify English /r/ and /l/. III. Long-term retention of new phonetic categories. *J. Acoust. Soc. Am*. 96, 2076–2087.

[DAY 3: Poster Session 3]

Acoustic characteristics and variation of clicks in the endangered language Nǀuu

Carina Marquard, Oliver Niebuhr, & Alena Witzlack-Makarevich

Dept. of General Linguistics, ISFAS, Kiel University, Germany

carina.marquard@web.de;{niebuhr;awitzlack}@isfas.uni-kiel.de

Clicks, i.e. “velarically initiated, ingressive suction stops” (Ogden, 2009), are of growing interest in the speech sciences. The non-pulmonic stops have long been neglected, partly because they were predominantly associated with “exotic” African languages like !Xóõ, Khoekhoegowab, and Zulu, for which there is no strong phonetic-phonological research tradition. A new line of research showed that clicks are also anything but absent from European languages, although they occur only non-phonemically and with different speaker-specific frequencies. But even in this new line of research, whose importance has recently been stressed by a special issue on non-pulmomic sounds in JIPA (Simpson, 2013), studying clicks is mainly limited to functional analyses based on an auditory detection and description of the sounds.

Our study aims at supporting a complementary line of research on the instrumental measurement of clicks, as in Simpson (2007), Fuchs et al. (2007) and Fuchs & Rodgers (2013). With a stronger focus on comparative, context-related analyses of acoustic details, we are concerned with the following phonetic-phonological questions: (1) What are the relevant acoustic characteristics and dimensions that distinguish clicks at different places of articulation? (2) Do clicks vary along their relevant acoustic dimensions in different segmental and prosodic contexts? (3) In particular, given that clicks result from complex coordinated movements of two active articulators and are essentially characterized by their burst and fricative elements, are clicks also subject to speech reduction, just like pulmonic stop consonants?

It is obvious that all questions, but particularly (2) and (3), can best be investigated with a spontaneous speech corpus in which the clicks occur as predictable parts of the sound patterns of words, i.e. phonemically. For this reason, we ana-lysed a corpus of spontaneous Nǀuu, recorded during fieldwork studies of the third author. Nǀuu is the last living mem-ber of !Ui, and a moribund language currently spoken by less than 10 elderly individuals in West South Africa (Miller et al., 2009). Our developing speech sample includes by now 62 bilabial, dental, and palatoalveolar click phonemes that occurred at the onset of di- or trisyllabic phrase-final high-frequency nouns like /ʘoe/ (meat), /ǀaeki/ (woman), or /ǂoo/ (man). In these prosodically controlled contexts, we investigated the acoustic variation of clicks in the form of stress-level induced degrees of speech reduction. The level of sentence stress of the words containing the clicks was labeled perceptually by means of the 4-step Kiel PROLAB system. The clicks were analyzed in Praat with respect to (A) total duration, (B) number of release bursts and (C) temporal distance between them. Click-related fricative sections were measured in terms of (D) mean energy level, (E) lower spectral energy boundary, (F) number of spectral peaks, (G) frequency of spectral energy maximum, as well as (H) the center of gravity (I) and its skewness. Measurements (A) and (D)-(I) were also taken for a control sample of voiceless pulmonic stops in similar lexical and prosodic conditions.

Our statistical analyses of the measurements included one-way MANOVAs based on Click Type or Stress Level, as well as linear discriminant analyses with each of these factors. The MANOVA for Click Type yielded a highly significant main effect. Bilabial, dental, and palatoalveolar clicks differed most significantly in (A), as well as in the levels of (D), (E), and (H) in the corresponding fricative elements. Accordingly, the prediction of Click Type by the linear discriminant analysis was significantly above chance level with a top discrimination performance of 85% for the bilabial clicks and an average discrimination performance of 74.2%. As regards effects of sentence stress on the click patterns, results were overall less clear. There was a significant main effect of Stress Level on the measurements in the MANOVA, but with a smaller effect size than for Click Type. Stress Level differences mainly concerned (A)-(C), as well as (D) and (F) in the associated fricative elements. Likewise, the prediction performance of the linear discriminant analysis for Stress Level was still significant, but only with a top performance of less than 40%.

In summary, our results extend previous findings by showing, on the example of a fieldwork-based spontaneous speech corpus of the endangered language Nǀuu, that clicks have rich acoustic profiles. These profiles clearly separate different types of clicks along many temporal, spectral, and energy dimensions. We also provided initial evidence that clicks are subject to variation, such as stress-level induced speech reduction. However, especially in comparison to analogous acoustic and statistical analyses of our control sample, it seems that the reduction of clicks is limited in both rate and degree. That is, click reduction across all Nǀuu speakers often does not go as far as for pulmonic stop consonants, although we found rare cases in which clicks were reduced to approximants or even just changes in voice quality. Our findings are relevant for theories of speech reduction/variation in production and perception, including the H&H theory.

References

Fuchs, S., Koenig, L., Winkler, R. 2007. Weak clicks in German? *Proc. 16th ICPhS Saarbrücken*, 449-452. / Fuchs, S., Rodgers, B. 2013. Negative intraoral pressure in German: Evidence from an exploratory study. *J. Internat. Ph. Assoc. (JIPA)* 43, 321-338. / Miller, A. Brugman, J., Sands, B., Namaseb, L., Exter, M., Collins, C. 2009. Differences in airstream and posterior place of articulation among Nǀuu clicks. *JIPA 39*, 129-161 Ogden, R. 2009. *An Introduction to English Phonetics*. Edinburgh: EUP. / Simpson, A. 2007. Acoustic and auditory correlates of non-pulmonic sound production in German. *JIPA* 37, 173-182. / Simpson, A. 2013. Non-pulmonic sounds in European languages: Introduction. *JIPA* 43, 247-248.

Determining the representation of phonotactic restrictions with nonce words

Gillian Gallagher
New York University
ggillian@nyu.edu

A repetition and discrimination task with nonce words was conducted with 19 speakers of Cochabamba Quechua (CQ) to diagnose the representation of a categorical ordering restriction in the language, which prohibits medial ejectives in roots with initial plain stops, e.g., [rit'i] 'snow', *[pat'i]. Results suggest that the absence of roots with a plain-ejective combination has been learned and represented as an articulatory preference for ejectives at the left edge of the root.

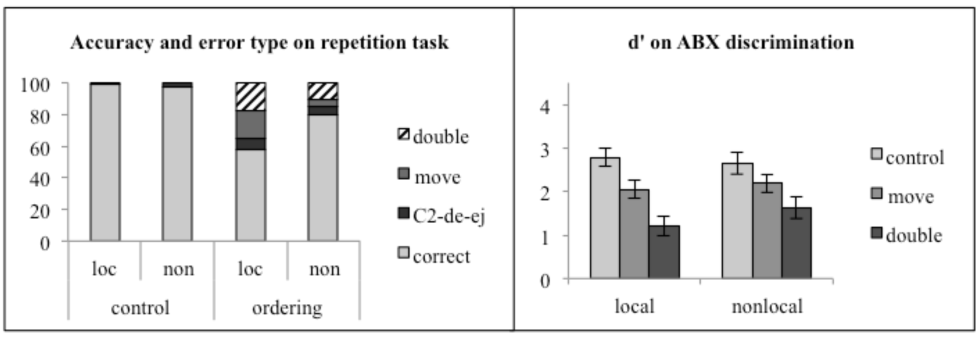
Repetition design: The target stimuli in the repetition task had an initial plain stop followed by an ejective in the adjacent syllable ([kap'i]) (local condition) or a non-adjacent syllable ([kamip'a]) (non-local condition). Control stimuli had phonotactically legal ejectives in the last syllable. Transcribed responses were coded for accuracy and type of error. Errors on the control category were always de-ejectivization of C_2, e.g., [sap'i] repeated as [sapi]. There were three types of errors on the ordering category: C_2 de-ejectivization ([kap'i] as [kapi]), movement ([kap'i] as [k'api]) and doubling ([kap'i] as [k'ap'i]). C_2 de-ejectivization and movement errors are phonotactic repairs, changing a phonotactically illegal target into a legal structure. Doubling, however, is not a repair, as roots with pairs of ejectives are also categorically absent in the language (*[k'ap'i]); this error, then, trades one unattested structure for another.

Repetition results: In both the local and non-local conditions, accuracy on targets that violate the ordering restriction is lower than accuracy on controls ($p < .001$). Accuracy on ordering restriction targets improves, however, from the local to non-local conditions ($p < .001$), suggesting that this restriction is distance sensitive. Errors on the ordering category are evenly split between repairs and non-repairs, in both the local and non-local condition, suggesting that errors are not soley driven by the phonotactic grammar, but also reflect phonetic difficulty. The improvement in accuracy in the non-local condition is primarily due to a reduction in the number of movement errors; though doubling errors are also slightly less frequent.

Discrimination: The discrimination task was an ABX task that compared the perceptibility of contrasts between three of the target-error pairs from the repetition task, in both local and non-local conditions, with the goal of determining the role of misperception in accountinf for errors. It is found that the contrast between a control target and a C_2 de-ejectivization error (e.g., nup'a-nupa) is the most discriminable ($p < .001$). The contrast between an ordering target and a movement error (e.g., kap'i-k'api) is more discriminable than the contrast between an ordering target and a doubling error (e.g., kap'i-k'ap'i) ($p < .001$). Performance on the local and non-local conditions does not differ. These results are fully consistent with the pattern of repetition errors in the non-local condition, but inconsistent with the pattern of repetition errors in the local condition. In the non-local condition of the repetition task, errors on controls are least common, and movement errors are less common than doubling errors for ordering targets; the frequency of errors is inversely correlated with the perceptibility of the contrast between the target and error on the discrimination task. In the local condition of the repetition task, however, doubling errors and movement errors are equally common for ordering targets, while the discrimination task shows that ordering targets are *less* confusable with a movement error than a doubling error. The rate of movement errors in the local condition of the repetition task is thus greater than would be expected by misperception alone.

The mismatch between perception and production for movement errors, as well as the sensitivity of these errors to locality, suggests the effects of an articulatorily substantive constraint. Additionally, doubling errors are consistent with an articulatory preference for ejectives at the left edge, and the non-repair status of these errors suggest they are not driven by a top-down preference for phonotactic wellformedness in the traditional sense.

Figure 1: Results of the repetition and ABX tasks. Repetition results show proportion of responses by accuracy and error type; discrimination results show d' by contrast type.



"Chilcotin flattening" revisited:
A phonetic investigation of Tsilhqut'in retraction effects

Sonya Bird
University of Victoria
sbird@uvic.ca

Tsilhqut'in (Chilcotin) is a Northern Athabaskan language spoken in Central Interior British Columbia by approximately 560 people. Its sound inventory is unique among Athabaskan languages in its size and complexity, sharing many features with neighbouring Salish languages. Of particular interest here, Tsilhqut'in contrasts plain vs. pharyngealized coronal consonants (e.g. /s/ vs. /sˤ/) and velar vs. uvular consonants (e.g. /k/ vs. /q/) (Krauss, 1975; Cook, 1993, 2013). The topic of this study is **vowel retraction effects** triggered by the pharyngealized coronal (Sˤ) and uvular (Q) consonants, as described in Cook (1993, 2013). Taken together, these effects provide an excellent window into the complexity of assimilation and harmony processes, even within a single language. However, while they have been widely cited in the phonological literature (Gick & Wilson, 2006; Hansson, 2010; Rose, 1996; and others), Cook's descriptions of them are based on his auditory impressions alone, and provide very little phonetic detail of the processes involved. The goal of this paper is therefore to verify his descriptions, based on auditory and acoustic analyses.

Recordings of illustrative words from Cook (1993 and 2013) were made with a single Tsilhqut'in speaker, a recognized language expert and one of Cook's own primary consultants. The vowels in these words (915 tokens total) were transcribed phonetically (auditory analysis) and vowel quality – the most salient acoustic correlate of retraction – was compared across different retracting and non-retracting environments (acoustic analysis: F1, F2, F3). Preliminary results reflect the complexity of the vowel retraction processes at hand, and also illustrate the importance of conducting phonetic work to confirm phonological descriptions. In terms of forward (rightward) retraction, findings by and large support Cook's observations: Q forward retraction only ever affects the immediately following vowel. Thus, we get /**ʁi**ti/ > [**ʁᵊi**ti] 'I slept' and /**ʁə**lməɬ/ > [**ʁʌ**lməɬ] 'it is rolling'. In contrast, Sˤ forward retraction affects the immediately following vowel and, if this vowel is reduced, the following vowel as well. Thus, we get /**sˤi**tin/ > [**sˤᶺi**tin] (*[**sˤᶺit**ʶᶺ**i**n]) 'I am sleeping' but /**sˤəɬti**n/ > [**sˤʌɬt**ʶᶺ**i**n] (*[**sˤʌ**ɬtin]) 'he is comatose'. These forms also illustrate that 1) retraction affects intervening consonants as well as vowels: /t/ in /sˤəɬ**t**in/ surfaces with secondary uvular frication [**t**ʶ]; and 2) the phonetic manifestation of retraction with /i/ is slightly but systematically different following Q vs. Sˤ: [ᵊ**i**] vs. [ᶺ**i**]; this likely reflects the difference between *uvularization* (with Q) and *pharyngealization* (with Sˤ). In terms of backward (leftward) retraction, results only partially support Cook's observations. As described by Cook, Q backward retraction only ever affects the immediately preceding vowel, as in /ʔæl**æχ**/ > [ʔæl**ɑχ**] (*[**ʔɑɬɑχ**]) 's/he is making it'. However, the current study shows that, even in this environment, retraction is not always clearly present, particularly with /i/, e.g. /n**iq**in/ > [n**iq**ˣᵊin]~[n**ɪq**ˣᵊin] (*[n**ɛq**ˣᵊin]) 'we paddled'. In terms of Sˤ backward retraction, results confirm Cook's finding that it occurs over larger domains, e.g. /**ʔæbələsˤ**/ > [**ʔʌbʌɬʌsˤ**]. However, here too retraction is variable: again, /i/ seems to resist retraction, e.g. /n**isˤ**dzˤun/ > [n**isˤ**dzˤõ]~[n**ɪsˤ**dzˤõ] (*[n**ɛsˤ**dzˤõ]) 'owl'. More importantly, contrary to Cook's description, Sˤ backward retraction never spreads past an intervening non-retracted consonant, e.g. /nægʷ**ənisˤ**tsˤəl/ with velar /gʷ/ > [nægw**ʌnɛsˤ**tsˤʌɬ] (*[n**ɑgwʌnɛsˤ**tsˤʌɬ]) 'fire has gone out'.

These findings suggest three things: 1) contrary to Cook (1993) but similar to the Interior Salish facts (e.g. Bessell 1998), **consonants** are affected by retraction as well as vowels – this is particularly evident with /t/ and /l/, which surfaced as [tʶ] and [ɫ] in retracted environments; 2) although Cook analyses both Q and Sˤ flattening as phonological processes, the differences in **domains** between Sˤ and Q effects suggest that while Sˤ flattening is indeed phonological, Q flattening is not; rather, it is a low-level phonetic process motivated purely by articulatory restrictions (Gick & Wilson 2006); and 3) based on the **variability** observed in the domains of backward-flattening (both with Q and Sˤ), these effects may be weakening over time, as in fact suggested by Cook (2013). The differences between the current findings and Cook's (1993, 2013) observations highlight the importance of conducting detailed phonetic fieldwork with a range of speakers, to fully understand sound-related processes in languages (including their evolution over time), and consequently to understand the implications of such processes for phonological theory.

References

Bessell, N. 1998. Local and non-local consonant-vowel interaction in Interior Salish. *Phonology 15*, 1-40. / Cook, E.-D. 1993. Chilcotin flattening and autosegmental phonology. *Lingua 91*, 149-74. / Cook, E.-D. 2013. *A Tsilhqút'in grammar*. Vancouver: UBC Press. / Gick, B., I. Wilson. 2006. / Hansson, G. 2010. *Consonant harmony: Long-distance interaction in phonology*. Berkeley: University of California Press. / Krauss, M. 1975. Chilcotin phonology: A descriptive and historical report with recommendations for a Chilcotin orthography. Unpublished. / Rose, S. 1996. Variable laryngeals and vowel lowering. *Phonology 13*: 73-117.

Vowel-to-vowel Coarticulation in Australian Languages: Place Matters

Simone Graetzer, Janet Fletcher and John Hajek

Phonetics Laboratory, University of Melbourne

{simone.graetzer; j.fletcher; j.hajek}@unimelb.edu.au

Coarticulation – the influence of one segment on another – is an important source of variability in the speech signal. As such, it has been argued that any model of speech motor control and of phonological organisation should be able to account for coarticulation patterns (e.g., Munhall, Kawato, & Vatikiotis-Bateson, 2000). In the present study, trans-consonantal anticipatory and carryover vowel-to-vowel coarticulation is investigated in three Australian indigenous languages, Burarra, Gupapuyngu and Warlpiri, in order to determine the magnitude of such coarticulation across various consonant places of articulation. It is well known that a number of factors can influence the extent of trans-consonantal coarticulation, including paradigmatic factors such as size and composition of phonological inventory and physiological factors such as degree of segmental articulatory constraint. An investigation of language-specific behaviour in this context is particularly important given that the majority of previous studies have focused on a small number of (mostly European) languages. There has been little or no study of the extent of this type of coarticulation in languages that have multiple (coronal) places of articulation, such as Australian languages, which are notable for their rich set of place of articulation contrasts (including up to four coronal contrasts). Conversely, these languages are also notable for their small vowel inventories. Most have only three contrastive vowel qualities and there is a high level of vowel allophony due to the effects of adjacent consonants. The evidence presented to date for Australian languages suggests that they have “consonant-salient” inventories and that they show strong resistance to the kinds of synchronic assimilatory processes that are observed in a range of other languages (Butcher, 2006). The extent of trans-consonantal vowel-to-vowel coarticulation may therefore also be modulated by the need to preserve the place cues of the intervening consonant. Experimental studies of other languages have shown there are place-dependent differences in the extent of coarticulation (e.g., Öhman, 1966; Recasens, 1984), so we would predict that this will also be the case in these Australian languages.

In this study, the participants were three speakers of each language recorded in field sites in three remote locations in Australia where the languages are still spoken. We analysed real $C_1V_1C_2V_2$ words where V was one of the vowels /i a u/, and C_2 was a bilabial, alveolar, retroflex, (alveo-)palatal or velar stop. An example of a carryover comparison is /Cata/~/Cuta/, where the flanking vowel, V_1, varies between /a/ and /u/ and the target vowel, V_2, is /a/. Our analysis of F1, F2 and F3 variation shows that in these languages close vowels exert a degree of vowel-to-vowel coarticulation, more so than non-close vowels. We also demonstrate, as predicted, that the place of articulation of the intervening consonant modulates (anticipatory and carryover) vowel-to-vowel coarticulation. This intervening consonant appears to affect target vowel spectra more systematically than does the flanking vowel, especially when the place of articulation is coronal. Hence, it could be argued that coarticulation is limited primarily by intervening consonantal rather than flanking vocalic output constraints, reflecting the organisation of articulatory programming in these languages (see, e.g., Russian: Öhman, 1966; Iskarous & Kavitskaya, 2010). This finding is broadly consistent with Butcher’s (2006) ‘place of articulation imperative’, which proposes that the need to maintain phonological consonant place distinctions in the face of potential synchronic coarticulatory pressures is a dominant constraint in Australian languages. One of those competing pressures is degree of articulatory (tongue dorsum) constraint (see, e.g., Recasens, 1984), which we argue is a major factor differentiating close and non-close vowel and coronal and non-coronal consonant behaviour in this context. However, the realisation of vowel-to-vowel coarticulation in these data is acoustically variable, as found in studies of languages such as English (e.g., Brancazio & Fowler, 1998). In addition, there is evidence of inter-language and inter-speaker variability. To the authors’ knowledge, this study is the first to investigate vowel-to-vowel coarticulation in Australian languages.

References

Brancazio, L., & Fowler, C. A. 1998. On the relevance of locus equations for production and perception of stop consonants. *Percept. Psychophys.*, 60, 24-50. / Butcher, A. R. 2006. Australian aboriginal languages: consonant-salient phonologies and the 'place-of-articulation' imperative. In: Harrington, J. M., Tabain, M. (eds.), *Speech Production: Models, Phonetic Processes and Techniques*. New York: Psychology Press, 187-210. / Iskarous, K., & Kavitskaya, D. 2010. The interaction between contrast, prosody, and coarticulation in structuring phonetic variability. *J. Phon.* 38, 625–639. / Munhall, K. G., Kawato, M., & Vatikiotis-Bateson, E. (2000). Coarticulation and physical models of speech production. In: Broe, M., Pierrehumbert, J. (eds.), *Papers in Lab. Phon. V. Acquisition and the Lexicon.* Cambridge: Cambridge University Press, 9-28. / Öhman, S. E. G. 1966. Coarticulation in VCV utterances: spectrographic measurements. *J. Acoust. Soc. Am*. 39, 151-168. / Recasens, D. 1984. Vowel-to-vowel coarticulation in Catalan VCV sequences. *J. Acoust. Soc. Am*. 76, 1624-1635.

Schwa in Tashlhiyt Berber in voiceless environments

Timo B. Roettger[1], Rachid Ridouane[2] & Martine Grice[1]
[1]IfL Phonetik, University of Cologne;
[2]Laboratoire de Phonétique et Phonologie (UMR 7018) CNRS/Sorbonne Nouvelle
{timo.roettger; martine.grice}@uni-koeln.de; rachid.ridouane@univ-paris3.fr

The status of *schwa* in Tashlhiyt Berber has long been controversial. Some have analysed it as a bona fide vowel, occupying the syllable nucleus (e.g. Coleman 2001), others see it as an intrusive element, with a distribution predictable from the consonantal environment: non-homorganic clusters with voicing specified for at least one consonant (Ridouane & Fougeron 2011). In the latter case, schwa is interpreted as a phonetic artifact. However, schwa has been attested in entirely voiceless words (Ridouane 2008, Roettger et al 2012, 2013), although this was in utterances produced by Paris-based speakers, making it difficult to rule out an influence of French.

This study investigates schwa in voiceless environments with speakers recorded in the field, in Agadir (Morocco). The aim is to determine whether phrase position conditions the distribution and duration of schwa. Ten speakers read a variety of contextualized sentence types varying in the position of inserted target words (phrase final vs. phrase medial). In the present study we focus on voiceless target words (e.g. /tʃtft/ 'you crushed'). They either appeared phrase finally (e.g. <Inna TARGET> 'he said TARGET.'), or phrase medially, e.g. followed by <abadan> 'always' (e.g. <Inna TARGET abadan> 'he said TARGET always.').

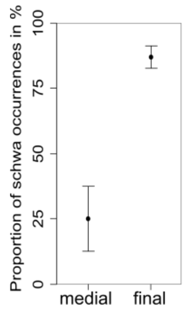
We counted instances of schwa and measured their duration. Preliminary results for four speakers (1M, 3F) are displayed in Fig. 1 and 2. Overall, there was considerable variation in the position of schwa in the target words, both across and within speakers, with it occurring either interconsonantally (e.g., [tʃtəft]) or at the end of the word (e.g. [tʃtftə]). Schwas were more frequent when the target word was phrase final (87% of cases, n=192), as opposed to medial (25% of cases, n=480), and exhibited a greater mean duration in phrase final target words (91ms) as opposed to phrase medial words (58ms). Thus, given the influence of the position of the word in the phrase in determining its presence and duration, it is difficult to argue for all schwa occurrences as mere phonetic artifacts. Furthermore, since speakers did not use French on a daily basis, we are able to validate earlier findings based on speakers from Paris, ruling out a transfer from French.

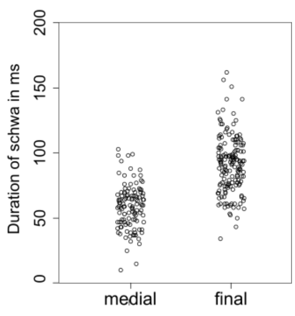
The question arises as to why schwas appear in these contexts. Comparing fully voiced and entirely voiceless target words (cf. Fig 3), we can observe a tonal event on the former, and, typically, a schwa with a pitch movement on the latter. The pitch on the schwa reflects salient aspects of the fully voiced contour. However, if the entirely voiceless target is not final, there is often a prominent pitch peak on the final word in this phrase instead. There were fewer and shorter schwas in these (non-final) cases. We might thus conclude that speakers exploit schwa to realize tonal events if the phonetic opportunities for the execution of intonational pitch movements are limited, as is the case when the target word is completely voiceless.

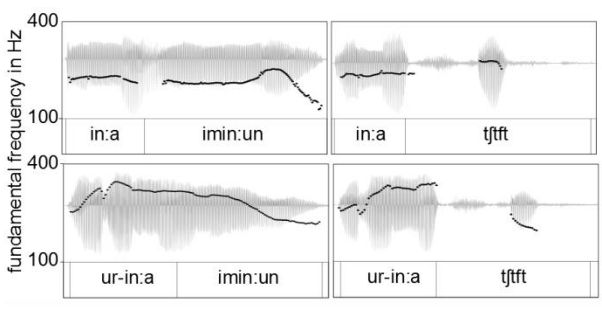
Figure 1 (left): Means (and SEs) of occurrences of schwa as a function of the position of the target word in the phrase, either medial or final.

Figure 2 (middle): Strip chart of duration of schwa as a function of the position of the target word in the phrase, either medial or final.

Figure 3 (right): Representative F0-contours and waveforms for 'he said TARGET' and 'he didn't say TARGET'. Left: fully voiced target (/imin:un/ 'your mouths'); Right: voiceless target (/tʃtft/ 'you crushed'), which surfaces with a schwa ([tʃtəft]).







References

Coleman, J. 2001. The phonetics and phonology of Tashlhiyt Berber syllabic consonants. *Transactions of the Philological Society* **99**. 29–64. / Roettger, T.B., Ridouane, R. Grice, M. 2013. Phonetic alignment and phonological association in Tashlhiyt Berber. Proc 21st ICA. / Roettger, T.B., Ridouane, R. Grice, M. 2012. Sonority and syllable weight determine tonal association in Tashlhiyt. Proc. 6th Speech Prosody. / Ridouane, R. 2008. Syllables without vowels: phonetic and phonological evidence from Tashlhiyt Berber. *Phonology* **25**. 321–359. / Ridouane, R., Fougeron, C. 2011. Schwa elements in Tashlhiyt word-initial clusters. *Journal of Laboratory Phonology* **2**. 1–26.

Getting to the root of the problem: An ultrasound investigation of 'Advanced Tongue Root' in Lopit

Rosey Billington

Phonetics Laboratory, University of Melbourne

rbil@unimelb.edu.au

Many languages of Africa (and some elsewhere) are described as having vowel inventories in which the phonological feature 'Advanced Tongue Root' (ATR) distinguishes vowels of a similar height, and also forms the basis of a vowel harmony system. Vowels are typically classed as either [+ATR] or [-ATR]. Although the phonological feature is widely attested, phonetic explorations of ATR remain limited. Research to date suggests that while ATR vowel systems may be phonologically similar across languages, a range of possible articulatory and acoustic correlates exists (Casali 2008). There are also signs that the correlates of ATR may be less uniform in the Nilo-Saharan languages of East Africa than in the Niger-Congo languages of West Africa, but studies of ATR in Nilo-Saharan languages remain few in number. To contribute to the understanding of ATR in Nilo-Saharan languages, I present acoustic and articulatory data on Lopit, an un(der)described Eastern Nilotic minority language of South Sudan. I focus in particular on what Ultrasound Tongue Imaging (UTI) reveals about the mapping between the phonology and the phonetics of ATR in this language.

In terms of articulation, the phonological feature ATR is conventionally held to correlate with movement of the tongue root as a consequence of pharyngeal expansion, e.g. as discussed by Lindau (1979) with reference to many radiographic studies of Niger-Congo languages and by Tiede (1996), who used Magnetic Resonance Imaging (MRI) to study vowels in Akan (also Niger-Congo). Recent investigations using UTI provide supporting evidence for tongue root advancement as the gestural correlate of [+ATR] in the Niger-Congo languages Dagbani (Hudu, Miller & Pulleyblank 2009) and Yoruba (Allen, Pulleyblank & Ajíbóyè 2013). However, for Nilo-Saharan languages, the findings appear to be less consistent. Jacobson (1978) observes that in Southern Nilotic DhoLuo, either tongue height or tongue root may be used to distinguish the vowel category, and that speakers vary in which strategy they prefer. For the Eastern Nilotic language Ateso, Lindau, Jacobson and Ladefoged (1972) find that the main difference between the two classes of vowels is one of tongue body height. These results from radiographic studies lead Jacobson (1978) to suggest that East African languages may exhibit greater variability or flexibility in the gestural correlates of ATR than West African languages. To investigate this, articulatory data for more languages is required.

In Lopit, the presence of an ATR contrast has been noted in existing wordlists and accompanying phonology sketches (Vossen 1982; Turner 2001), but further research on the language has only recently begun. Language documentation in linguistically diverse South Sudan has been hampered by the many decades of civil war in the former Sudan, as well as current political instability in the recently independent South. Millions of people have died during wartime, and millions more have been displaced, and the extent to which this population upheaval has affected minority languages is not yet known. However, the existence of vibrant diaspora groups provides opportunities to study otherwise inaccessible minority languages. The present study is part of a wider project documenting the phonetics and phonology of Lopit through work with a small community (~ 8 families) of Lopit speakers in Melbourne, Australia. I present phonological evidence for an ATR-type system in Lopit, then discuss acoustic results for four male speakers producing [+ATR] vowels /i, e, o, u/ and [-ATR] vowels /ɪ, ɛ, ɔ, ʊ/ in a 100-item wordlist. I then present the results of mid-sagittal UTI for two speakers, using a subset of the wordlist items. Results show that the acoustic correlates of ATR in Lopit are broadly similar to those observed for ATR in other languages, but that there are some notable differences compared to other Nilotic languages. UTI data reveal that while the lingual gestures used in the articulation of [+/-ATR] vowels fit with basic assumptions about 'Advanced Tongue Root', they differ from what might be expected on the basis of data for related languages such as Ateso. These findings support the idea that ATR is perhaps best considered an abstract 'cover feature' for a range of articulatory and acoustic correlates (e.g. Hyman 1988), and reinforce the need for further instrumental investigations of such vowel systems, particularly in under-documented language families.

References

Allen, B., Pulleyblank, D. & Ajíbóyè, O. 2013. Articulatory mapping of Yoruba vowels: an ultrasound study. *Phonology* 30, 183-210./ Casali, R. F. 2008. ATR Harmony in African languages. *Language and Linguistics Compass* 2, 496-549./ Hudu, F., Miller, M. & Pulleyblank, D. 2009. Ultrasound imaging and theories of tongue root phenomena in African languages. In Austin, P. K., Bond, O., Charette, M., Nathan, D. & Sells, P. (eds.), *Proc. LDLT*, 153-163. / Hyman, L. M. 1988. Underspecification and vowel height transfer in Esimbi. *Phonology* 5, 25-73. / Jacobson, L. C. (1978). DhoLuo Vowel Harmony: A phonetic investigation. *UCLA Working Papers in Phonetics,* 43, 1-122. / Linda, M. 1979. The feature 'expanded'. *J. Phonetics* 7, 163-176. / Lindau, M., Jacobson, L. C. & Ladefoged, P. 1972. The feature Advanced Tongue Root. *UCLA Working Papers in Phonetics* 22, 76-94. / Tiede, M. K. 1996. An MRI-based study of pharyngeal volume contrasts in Akan and English. *J. Phonetics* 24, 399-421. / Turner, D. 2001. Lopit phonology. Unpublished manuscript, SIL-Sudan. / Vossen, R. 1982. *The Eastern Nilotes: Linguistic and Historical Reconstructions*. Berlin: Dietrich Reimer Verlag.

Evidence of mismatch between tonal production and perception in Karbi

Amos Teo and Linda Konnerth

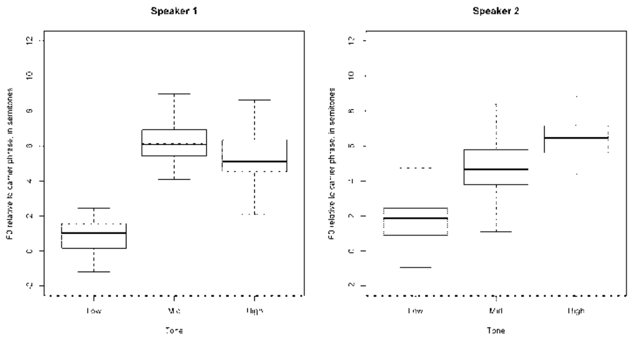
University of Oregon

ateo@uoregon.edu, lkonnert@uoregon.edu

This paper provides preliminary evidence for a mismatch between the production and perception of tones in Karbi, a Tibeto-Burman language spoken in Assam. It represents one of the first phonetic studies of an under-described minority language of India.

In a previous description of Karbi, Grüßner (1978) described three tones: low, mid and high, which are all distinguished by pitch height. He further noted that mid tones in word-final position are accompanied by syllable-final glottalization. However, given technical difficulties associated with doing fieldwork in the area at the time, this description was necessarily based on researcher-centered impressionistic evidence. More recently, Anonymous (submitted) conducted an acoustic study of monosyllabic stems in carrier phrases and found evidence to support Grüßner's claim that the low and high tones in Karbi are distinguished in production by pitch height. On the other hand, of the two speakers included in the study, only Speaker 2 produced pitch patterns that significantly distinguished a low, a mid and a high tone (see Figure 1). Both speakers did produce word-final mid tone stems with syllable-final glottalization, as described by Grüßner, but this was lost when suffixes were added to the stem.

Figure 2: Box-and-whisker plots comparing two Karbi speakers' productions of pitch relative to a carrier phrase at the vowel mid-point of monosyllabic stems



An initial hypothesis was that the contrast between the mid and high tones was neutralized on suffixed stems for Speaker 1 but not Speaker 2, who continued to distinguish them by pitch height in production. This hypothesis was tested in a perception experiment involving six listeners who were asked to identify both word-final and suffixed monosyllabic stems, as spoken in carrier phrases by Speakers 1 and 2. In addition, Speaker 2 listened to his own recordings to identify the stems.

The results of the perception study showed that, contrary to expectations, the difference in pitch height produced by Speaker 2 was not found to assist listeners in differentiating mid and high tones on suffixed stems. Even more strikingly, it was found that Speaker 2 himself was unable to reliably identify suffixed stems, despite producing a significant difference in pitch height in his own recordings. Given that listeners were still generally able to identify monosyllabic stems in word-final position as produced by both Speakers 1 and 2, it is likely that syllable-final glottalization, not contrastive pitch height, is the main perceptual cue for what has been called the 'mid tone' in Karbi.

The significance of these findings, particularly the mismatch between tone production and perception by Speaker 2, will be discussed in the talk. We consider the possibility of a near-merger situation (as per Labov et al., 1991) in Karbi, whereby pitch is not a reliable perceptual cue for the mid-high distinction within the language community, even for speakers who produce the distinction, perhaps because these speakers are exposed to speech where the distinction is not made. In addition, these findings raise the question of what it means to continue referring to Karbi as a 'tone' language.

References

Anonymous. Acoustic-statistical and perceptual investigations of Karbi tones: A peculiar case of incomplete neutralisation of F_0. In *North East Indian Linguistics.* Forthcoming. / Grüßner, K.-H. 1978. *Arleng Alam, die Sprache der Mikir: Grammatik und Texte*. (Beiträge zur Südasienforschung, Bd. 39). Wiesbaden, Franz Steiner Verlag. / Labov, W., Karen, M., Miller, C. 1991. Near-mergers and the suspension of phonemic contrast. *Language Variation and Change*. 3, 33-74.

Inserted vowel articulation in Scottish Gaelic: a preliminary report

Diana Archangeli#, Sam Johnston*, Jae-Hyun Sung*, Muriel Fisher*, Michael Hammond*, Andrew Carnie**

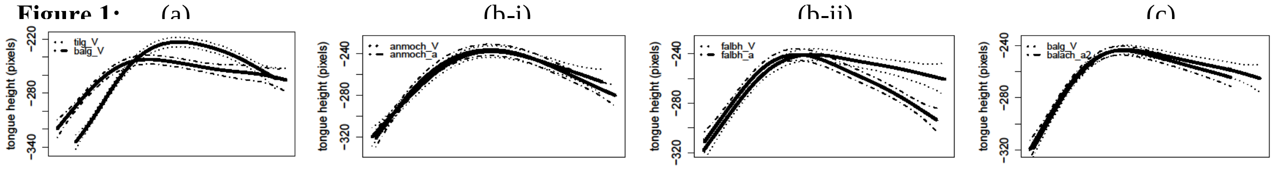
University of Arizona*, University of Hong Kong#

scsjohnston@email.arizona.edu, darchang@hku.hk, jhsung@email.arizona.edu, murielf@email.arizona.edu, carnie@email.arizona.edu

In Scottish Gaelic, vowels that appear between a sonorant and a following non-homorganic voiced obstruent are of two types. Compare lexically disyllabic *ballag* [balak] 'skull' with *balg* [balak] 'blister'. In words like *balg*, the second vowel is inserted, either historically (Borgstrøm, 1940) or synchronically (Clements, 1986; Bosch 1991; Bosch & de Jong, 1997; Smith 1999), and either phonetically or phonologically (Hall, 2006). The inserted vowel appears to be a copy of the preceding vowel, though other vowel qualities are sometimes shown (e.g. Sagey 1986, Gillies 1993); psycholinguistic studies distinguish the inserted vowel from non-inserted vowels (Hammond et al. to appear).

Our goal is to explore the articulation of the inserted vowel to better understand the nature of the vowel quality, and to test the phonetic vs. phonological hypotheses. Data were collected from 26 speakers of Scottish Gaelic, all of whom spoke Scottish Gaelic primarily through young childhood and still use the language regularly today. In this preliminary report, data from one subject is analyzed to answer three questions. (a) *Are inserted vowels in different words articulated with different tongue positions?* If inserted vowels copy the preceding vowel, there should be variation if the preceding vowel is different. (b) *Is the inserted vowel of a given word articulated with the same tongue position as the pre-inserted vowel in the same word?* The vowel copy hypothesis predicts that the two vowel articulations will be quite similar. (c) *Is the inserted vowel articulated with the same tongue position as a matched non-inserted vowel in a (near) minimal pair?* The phonological hypothesis predicts categorical similarity; the phonetic hypothesis predicts the possibility of gradient dissimilarity.

Figure 1: (a) (b-i) (b-ii) (c)



Representative tongue contours were compared using SSANOVA (Davidson, 2006) (Figure 1). In words matched for the surrounding consonants, (a): *tilVg* and *balVg*, inserted vowels (*V*) are quite different from each other suggesting that these vowels acquire their quality from some source other than neutral tongue position and/or coarticulation. Comparing the pre-inserted vowel with the inserted vowel, (b-i): *anVmoch* and (b-ii): *falVbh*, shows similarity in (b-i) but differences in (b-ii). This suggests that copy cannot be the sole source for quality of the inserted vowel. Comparing the inserted vowel with a non-inserted vowel in (near) minimal pairs, (c): *balVg* and *balach*, reveals almost complete overlap between the tongue contour for an inserted vowel and its paired non-inserted vowel, consistent with the two being phonologically identical, suggesting that the quality of the inserted vowel is due to the preceding vowel, and that the variation seen in cases like (b-ii) is the effect of coarticulation to the adjacent consonants. Closer examination of the (b-ii) cases reveals that (i) when the intervening consonant is a palatal, the inserted vowel is higher or more fronted (or both) than the preceding vowel; (ii) when the intervening consonant is an [l], the tongue tip is raised in the inserted vowel; and (iii) when the intervening consonant is [r], the tongue body is lower in the inserted vowel. We tentatively conclude that the inserted vowel is a phonological copy of the preceding vowel, with observed variation due to coarticulation, consistent with the phonological analyses of Sagey (1986), Halle, Vaux and Wolfe (2000).

References

Borgstrøm, C.H.J. 1940. *A Linguistic Survey of the Gaelic Dialects of Scotland*. Oslo, Norway: H. Aschehoug and Co. / Bosch, A. 1991. *Phonotactics at the Level of the Phonological Word*. Ph.D. thesis, U. of Chicago. / Bosch, A. & de Jong, K. 1997. The prosody of Barra Gaelic epenthetic vowels. *Studies in the Linguistic Sciences* **27**:1–15. / Clements, G.N. 1986. Syllabification and epenthesis in the Barra dialect of Gaelic. In K. Bogers, H. van der Hulst & M. Mous (eds.), *The Phonological Representation of Suprasegmentals.* Dordrecht: Foris Publications, 317–336. / Davidson, L. 2006. Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *J. Acoust. Soc. Am.* **120:**407-415. / Gillies, W. 1993. Scottish Gaelic. In *The Celtic Languages*, ed. by Martin Ball, James Fife, Routledge, 145-227. / Halle, M., B. Vaux, & A. Wolfe. 2000. On feature spreading and the representation of place of articulation. *Linguistic Inquiry* 31:387-444. / Hammond, M., N. Warner, A. Davis, A. Carnie, D. Archangeli M. Fisher. To appear. Vowel insertion in Scottish Gaelic. *Phonology*. / Ladefoged, P., Ladefoged, J., Turk, A., Hind, K., & Skilton, J. 1998. Phonetic structures of Scottish Gaelic. *J. of the IPA* **28**:1–41. /Sagey, E. 1986. The representation of features and relations in non-linear phonology, Ph.D. thesis, MIT. / Smith, N. 1999. A preliminary account of some aspects of Leurbost Gaelic syllable structure. In Harry van der Hulst & Nancy A. Ritter (eds.), *The Syllable:Views and Facts*. Berlin and New York: Mouton de Gruyter, 577–630.

Tonal Polarity in Xitsonga as a tonal artifact: an acoustic analysis

Siri Gjersøe & Seunghun J. Lee

Humboldt University & Central Connecticut State University

sirigje@gmail.com & juliolee@gmail.com

SYNOPSIS: Tonal polarity has been studied from phonological perspectives, but in this talk we report an acoustic study of tonal polarity in Xitsonga, a southern Bantu language spoken in South Africa. Xitsonga tonal polarity in additive plural prefixes is first reported in Cole-Beuchat (1959): a plural prefix is high tone (H) when added before a low-tone (L) initial root ([mí-làwù] 'CL4-law', H-LL), but it is L when added before an H initial root ([mà-trónkò] 'CL6-prison', L-HL). This talk analyzes tonal polarity patterns of two Xitsonga speakers and gives quantitative evidence against tonal polarity being at work in Xitsonga. A phonetic analysis suggests that what is considered to be tonal polarity should rather be characterized by an effect of the carrier sentence and root-initial consonants.

PREVIOUS PHONOLOGICAL ANALYSIS: While tonal polarity was first reported in additive plural prefixes, Lee (2013) observes that the other type of plural prefixes that substitute for singular prefixes do not display tonal polarity ([mú-gájó] 'CL3-mealie meal' H-HH, and [mí-gájó] 'CL4-mealie meal', H-HH). The asymmetry between the two types of plural prefixes is analyzed with a ranking in which the OCP-TONE constraint is outranked by the UNIFORMITY-TONE constraint, which is violated when the singular base and the plural stem have different tonal specifications.

METHOD: Acoustic data of two Xitsonga speakers was collected in Mhinga, South Africa. The speakers, in their 20's, produced target words in the L-toned carrier sentence *nì tìrhìsà* X *kàn'wè* 'I use X (=target) again'. After recording singular forms in the carrier sentence, they recorded plural forms based on the prompt "What do you say if you have multiple Xs?" Recordings were made with a sampling rate of 44.1 KHz, 16-bit into a Zoom H4 recorder using a Shure head-worn unidirectional microphone. 145 nouns that reportedly have a prefix with tonal polarity are analyzed.

RESULTS AND DISCUSSION: The results of the analysis show that tonal polarity only occurs in one direction when a root vowel has an H tone. If the samples displayed true tonal polarity in both directions, we expect that a monosyllabic root vowel (RV) with a monosyllabic additive prefix (ADD) should show a negative correlation in which the root vowel with a low pitch is preceded by the additive prefix with a high pitch. The figures 1 and 2, however, display a (weak) positive correlation, suggesting that the prefix vowel has a low pitch irrespective of the pitch of the root vowel. (Dots stand for lexical H tone; triangles stand for lexical L tone.) Noun classes, the quality of the root vowel and gender did not show significant effects.

Figure 1: Male speaker (RV for y-axis, ADD for x-axis)

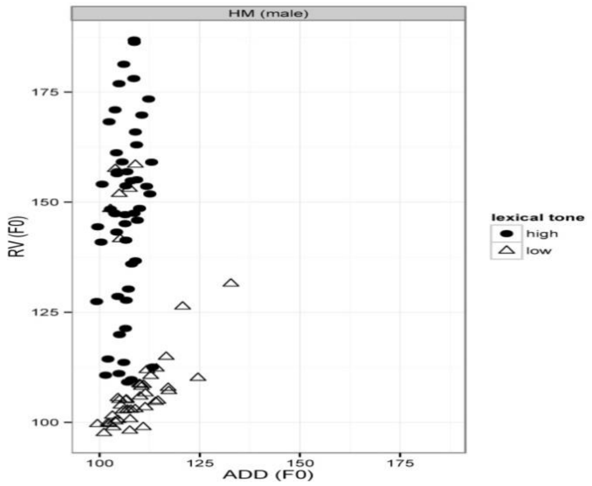
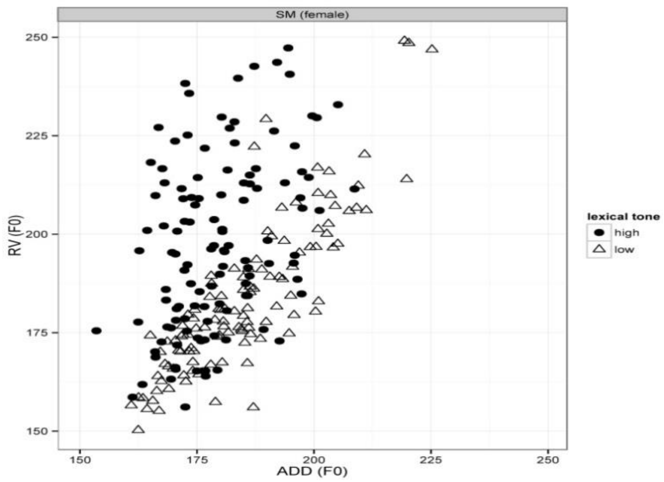


Figure 2: Female speaker (RV for y-axis, ADD for x-axis)



We suggest that the ADD prefix is L-toned and the previously attested H-L tonal polarity pattern is not really phonological but rather an artifact of various phonetic factors; (i) a high vowel of an H-toned ADD raises the pitch, and (ii) an initial pitch-lowering consonant (a depressor, cf. Lee 2009) of a lexical L-toned RV lowers the pitch. There were 6,9 % of the nouns that attested a tendency of a H-L pattern. For this data, (i) and (ii) can fully account for the H-L contrast. The co-articulation effect by depressors is also found when a lexical H-toned RV is preceded by an L-toned ADD; there were 28% of these nouns that show a minimal pitch difference (less than 10 Hz), rather than the expected L-H contrast. These nouns (20 out of 25) have an initial depressor and can therefore also be accounted for by (i).

References Cole-Beuchat, P-D. 1959. Tonemorphology of the Tsonga noun. *African Studies* 18(3), 133-145. Lee, S. 2009. H tone, depressors and downstep in Tsonga. *Selected Proceedings of ACAL 38*. Cascadilla, 26-38. Lee, S. 2013. Tonal polarity and paradigm uniformity in the nominal paradigm of Xitsonga. *Studies in phonetics, phonology and morphology* 19(1), 107-123.

Is longer always better?
Neurolinguistic evidence for asymmetric processing of consonant duration

Sandra Kotzor, Adam Roberts, Allison Wetterlin & Aditi Lahiri
Language & Brain Laboratory, University of Oxford
sandra.kotzor@ling-phil.ox.ac.uk

Many languages use consonantal length to contrast words; e.g. in Bengali [pata] 'leaf' vs. [pat:a] 'whereabouts, location'. The main acoustic cue for differentiating length of consonants in Bengali is closure duration, geminates showing a closure duration of about twice that of singletons. But since duration contrasts are invariably relative, how do listeners process consonantal length when no other cues are available, and what does this tell us about the representation of duration in the lexicon? To investigate this question we used Bengali, a language with a productive word medial geminate-singleton contrast, and employed several methods: behavioural and evoked potential recordings in cross-modal lexical decision tasks as well as an MMN study.

In the MMN study, we used three pairs of words: a real word singleton with a geminate mispronunciation ([kena]/*[ken:a]), a real word geminate with a singleton mispronunciation ([g^{h}en:a]/*[g^{h}ena]) and a pair where neither geminate nor singleton are real words (*[ena]/*[en:a]). The overall results show a symmetrical discrimination effect for all pairs of stimuli containing a real word with both short and long deviants eliciting MMN responses of equal amplitude and latency. This indicates that any mismatch in duration is detected automatically.

In the lexical decision tasks, we used both form (fragment) and semantic priming with two sets of Bengali disyllabic words as auditory primes; lexical singletons with no geminate counterparts (e.g. [k^{h}ɔma] 'forgiveness'), and underlying geminates with no corresponding singleton word (e.g. [big:æn] 'science'). Nonword mispronounced primes were created by shortening or lengthening the medial consonant to create the corresponding (fake) geminate (*[k^{h}ɔma]) or singleton (*[bigæn]). In the form priming, we used fragment primes of two different lengths: CVC fragments which included the complete closure duration ([k^{h}ɔm] / [big:]) and CVC$_V$ fragments which contained an additional two glottal pulses of the following vowel ([k^{h}ɔm$_V$]/ [big:$_V$]). In the semantic priming, the primes were full words (cf. Table 1).

Table 1: Stimuli for semantic cross-modal priming experiments.

SHORT > LONG Experiments					
Task	Condition	Prime	Target	Priming (RT)	N400
SEMANTIC PRIMING	Singleton (word)	[k^{h}ɔma] 'forgiveness'	[marjona] 'forgiveness'	20ms*	low
	Geminate (nonword)	*[k^{h}ɔm:a]		32ms	low
LONG > SHORT Experiments					
Task	Condition	Prime	Target	Priming (RT)	N400
SEMANTIC PRIMING	Geminate (word)	[big:æn] 'science'	[gɔbeʃɔna] 'science'	15ms*	low
	Singleton (nonword)	*[bigæn]		3ms	high

The behavioural data for the fragment priming shows significant priming effects for both nonword conditions but the difference in the degree of priming of nonword and real words was only significant in the tasks where a geminate real word was primed by a singleton nonword fragment. The singleton nonword fragment facilitated reactions to the geminate real word target significantly less, while in the task with geminate nonwords facilitation was equal. This tendency surfaced as an asymmetric pattern of priming in subsequent semantic priming tasks and the effects are significant in both behavioural and ERP data (reflected by lower mean N400 amplitudes and faster reaction times). Nonword mispronunciations resulted in priming of the semantically related target but only when the real word was a singleton. In the case of geminate real words, singleton mispronunciations failed to activate the real word and thus its semantic associates. This asymmetry shows that although the brain is highly sensitive to durational differences, a perfect match of consonantal length is not necessarily required for lexical activation.

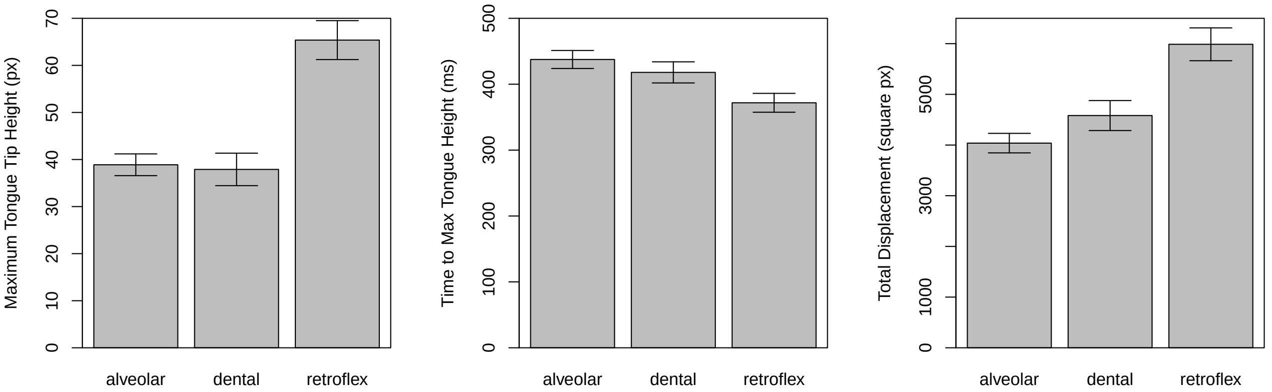
The overall findings in the present study indicate that this processing asymmetry stems from a difference in lexical specification: geminates are specified by an additional mora while singletons are not. A geminate mispronunciation subsumes the singleton real word representation because all other (featural) information is identical. However, when a singleton mispronunciation occurs in place of a geminate, the mora necessary for the match with a geminate representation cannot be generated from the duration in the acoustic signal and thus activation fails. Thus, full lexical access can be achieved through a mispronunciation only if its representation on the prosodic level does not mismatch with that of the corresponding real word.

The articulation of Kaytetye coronals

Susan Lin[1], Mark Harvey[2], Benjamin Davies[3], Myfany Turpin[4], Alison Ross[5], Katherine Demuth[3]
[1]University of California, Berkeley, [2]University of Newcastle, [3]Macquarie University,
[4]University of Queensland, [5]Northern Territories Department of Education
[1]`susanlin@berkeley.edu`

Kaytetye is an Australian language with minimally a four-way coronal place contrast – alveolar, retroflex, palatal, dental – across four manners, and is thus an excellent language for studying the articulation of these consonants. In this study, we examined ultrasound data collected from seven female Kaytetye speakers, targeting alveolars, dentals and retroflexes in the context /#a.ˈ_V/. Between four and six repetitions of each stimulus in the carrier phrase /aŋɡənə ŋə X/ (“Say X”) were elicited from each speaker.

Using the tongue contour at onset of the preceding /a/ as a basis for comparison, the maximum tongue tip HEIGHT in pixels (px) and TIME from vowel onset to maximum tongue tip height in milliseconds (ms) were calculated. Additionally, the total lingual DISPLACEMENT (px^2) from vowel onset to target constriction was calculated. This value represents the overall movement of the whole tongue, not localized to the tongue tip and blade. These data are illustrated in Figure 1a-1c.



(a) Max tongue tip height (px). (b) Time of max tongue height (ms). (c) Displacement (px^2).

Figure 1: Measures taken, by consonant place of articulation.

A series of linear mixed effects models with SUBJECT as a random factor were run. The data showed that tongue HEIGHT was significantly greater in retroflex than dental ($\beta = 27.1, t = 11.8, p < 0.0001$) or alveolar ($\beta = 26.2, t = 12.5, p < 0.0001$) consonants. Furthermore, speakers took less TIME to reach these targets in production of retroflex compared to alveolar ($\beta = -62.2$, $t = -5.8$, $p < 0.0001$) and dental ($\beta = -47.3$, $t = -4.0$, $p < 0.0001$) consonants. No significant difference in HEIGHT or TIME between dental and alveolar consonants was found. However, production of alveolar consonants resulted less total lingual DISPLACEMENT than dental consonants ($\beta = -583$, $t = 3.4$, $p = 0.0007$). Retroflex consonants also demonstrated more DISPLACEMENT than both alveolar ($\beta = -1953$, $t = 11.1$, $p < 0.0001$) and dental ($\beta = 1370$, $t = 7.1$, $p < 0.0001$) consonants.

Dentals and alveolars differed from each other only in terms of total lingual displacement, while tongue height and time to target were not significantly different. This corroborates our qualitative observation that the tongue body depresses during production of dentals, which promotes extension of the tongue tip and blade for formation of the dental constriction. In contrast, Kaytetye retroflex consonants differed from the others along several dimensions. They were produced with significantly more raised tongue tip and blade than dental and alveolar consonants, and the amount of time required to reach articulatory targets was significantly shorter. These facts are consistent with an account of retroflexion in which the articulatory requirements of sublaminal contact with the alveolar ridge are substantial, facilitated by early initialization and execution. This is further consistent with retroflex harmony in Kaytetye (Harvey, 2011) being due, at least in part, to anticipatory coarticulation (Walker et al., 2008).

References

Harvey, M. (2011). Prepalatals in Arandic. *Australian Journal of Linguistics*, 31(1):79–110.

Walker, R., Byrd, D., and Mpiranya, F. (2008). An articulatory view of Kinyarwanda coronal harmony. *Phonology*, 25:499–535.

Cross dialectal vowel spaces in Greek

Baltazani, M.[a,b]*, Kainada, E.* [a]*, Nikolaidis, K*[d]*., Sfakianaki, A.*[d]*, Lengeris, A.*[c,d]*, Tsiartsoni, E.*[d]*, Papazachariou, D.*[e]*, Giakoumelou, M.*[e]

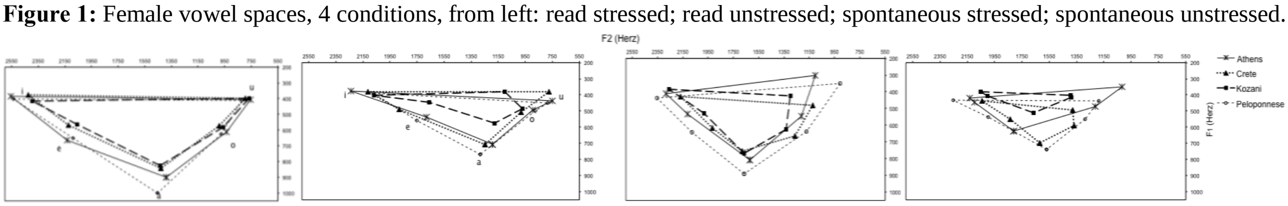
[a]Univ. of Ioannina, [b]Univ. of Oxford, [c]Univ. of Crete, [d]Artistotle Univ. of Thessaloniki, [e]Univ. of Patras
mbaltaz@cc.uoi.gr, ekainada@cc.uoi.gr, knicol@enl.auth.gr, asfakian@enl.auth.gr, lengeris@enl.auth.gr, eltsiart@gmail.com, papaz@upatras.gr, mgiakoumelo@upnet.gr

Cross-dialectal/linguistic surveys on vocalic systems allow for a better understanding of the possible organization of vocalic systems. Across languages, 75% of vowel inventories contain the same specific five to seven vowels, most often /i, e, a, o, u/ for five-vowels (Maddieson, 1984). Such proclivities have given rise to well-known hypotheses on the forces shaping vowel spaces: Dispersion Theory postulates that the distance among vowels serves to reduce confusion and increase perceptual contrast (Liljencrants and Lindblom, 1972; Lindblom, 1986); the Quantal Theory of speech (Stevens, 1989) suggests there are regions of stability in the phonetic space, in particular for the point vowels /i/, /a/, /u/ which are predicted to remain in the same positions across languages, regardless of vocalic inventory size. Our paper presents evidence suggesting a more complex picture than either of these hypotheses suggest.

Our data arise from a larger project (Vocalect), which contributes, for the first time, a comprehensive description and analysis of the vocalic systems of several Modern Greek (MG) regional dialects, most of which are endangered, at a phonetic and phonological level. One of the project's ultimate goals is the contribution to the creation of a dialect atlas for Greek, a resource so far missing (cf. Trudgill, 2003). We martial acoustic, articulatory (EPG) and perceptual analyses of vowel data collected through extensive fieldwork. Here we present results from a subset of our corpus: the vocalic system of four dialectal areas of Greece (Macedonia, Peloponnese, Athens and Crete) based on controlled and spontaneous speech material from 40 and 8 speakers respectively, balanced for sex and dialect (3,600 tokens).

We investigate the durations, F1XF2 acoustic space and formant distances between adjacent vowels as a function of dialect, gender, speech style and stress. Our preliminary analysis of female speech shows an effect of all aforementioned factors (Fig 1): vowels in unstressed and spontaneous conditions occupy less space and have different distributions and distances from those in stressed and read conditions; Cretan and Macedonian vowel spaces are more compressed (area of 390 and 370 KHz respectively calculated using Heron's formula) than the Athenian and Peloponnesian (500 and 516 KHz respectively). Back vowels appear much fronter in conversational speech; vowels are not maximally dispersed; the point vowels (/i/, /a/, /u/) are in different positions and with dissimilar acoustic distances across dialects; adjacent vowels are not equidistant across dialects.

Figure 1: Female vowel spaces, 4 conditions, from left: read stressed; read unstressed; spontaneous stressed; spontaneous unstressed.



The initial picture emerging is intriguing and suggests that the explanation behind vocalic space structure is more complex than the aforementioned hypotheses. The variation observed may relate to several factors. Previous literature indicates that acoustic vowel targets can differ in languages/dialects with the same number of vowel contrasts due to phonological or historical reasons; thus frequency intervals between adjacent vowels can relate to dialect-dependent patterns rather than to universal ones (e.g., Bradlow, 1995, Recasens & Espinosa, 2006). Such patterns are seen in the MG dialects. For example, Macedonian is known for massive vowel deletions (Topintzi & Baltazani, 2012) which lead to more closed syllables and a greater degree of palatalization in its consonantal inventory, factors which can correlate with the observed shift in vocalic space (cf Bradlow, 1995). Differences may also occur due to context-dependent variability across dialects, such as consonantal inventories and inherent vowel durations in different rates of production.

Overall, our work showcases how indispensable cross-dialectal variation studies are for the deeper understanding of speech production variation. We will discuss dialect-specific sources of variability under the scope of current theories on i) vowel dispersion, ii) direction of unstressed phonetic reduction, and iii) sociolinguistic variation across genders.

References

Liljencrants, J., Lindblom, B. 1972. Numerical Simulation of Vowel Quality Systems: The Role of Perceptual Contrast. *Language* 48, 4: 839-862. / Lindblom, B. 1986. Phonetic universals in vowel systems. In: Ohala JJ, Jeager JJ (eds) *Experimental Phonology*. Orlando, FL: Academic Press, 13-44. / Maddieson, I. 1984. *Patterns of sound*. CUP. / Trudgill, P. 2003. Modern Greek dialects: a preliminary classification. *J. of Greek Linguistics* 4, 45-63. / Stevens, K.N. 1989. On the Quantal nature of speech. *J Phon* 17: 3-45. / Bradlow, A.R. 1995. A comparative acoustic study of English and Spanish vowels. *JASA*, 97(3), 1916-1924. / Recasens D., Espinosa, A. 2006. Dispersion and variability of Catalan vowels, *Speech Communication*, 48, 645-666. / Topintzi N & Baltazani, M. 2012. The acoustics of high-vowel loss in a Northern Greek dialect and typological implications. In *Consonant Clusters and Structural Complexity*, Hoole, P. et al. (eds.), de Gruyter, 373-402.

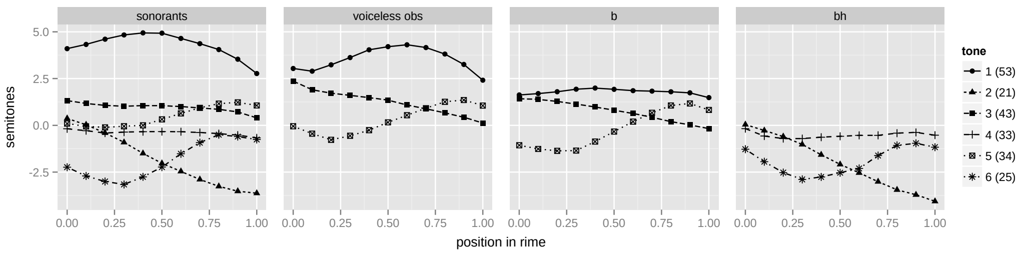
Laryngeal and tonal contrasts in the Tai dialect of Cao Bang

Pittayawat Pittayaporn[1] *& James Kirby*[2]
[1]Chulalongkorn University, [2]University of Edinburgh
[1]Pittayawat.P@chula.ac.th, [2]j.kirby@ed.ac.uk

Voice qualities have long been thought to play an important role in tonogenetic processes by mediating the transphonologization of segments into tones (Haudricourt, 1965). Breathiness, in particular, is thought to arise as a redundant phonetic cue to onset voicing, whose subsequent loss results in new tonal categories. However, it is still unclear what role breathiness plays at a stage when tonal contrasts are fully established, but the initial laryngeal contrasts have not yet been neutralized, in part because languages that preserve this state of affairs are extremely rare.

The Tai dialect of Cao Bang (CBT), spoken in northeastern Vietnam, is just such a language. In most Tai languages, a three-tone system split into six tones following the collapse of the voicing contrast. In CBT, voicing contrasts were neutralized following loss of aspiration in initial sonorants, but are still retained in initial obstruents, along with voice quality differences most other Tai languages have lost (Hoàng Văn Ma, 1997; Pittayaporn, 2009). As a result, the six-way system only applies to a subset of the laryngeal contrasts. While breathiness can be analyzed phonologically as a redundant feature of three of these tones, its phonetic status within the tone system is still unclear. In this paper, we present the first instrumental phonetic analysis of the language. Because tone and laryngeal properties crosscut one another in this complex system, we seek to understand how these contrasts are signaled phonetically.

Figure 2: F0 excursions on sonorant-initial, voiceless, and voiced initials. /bh/ initials are breathy-voiced.



Our acoustic analysis is based on recordings taken in Cao Bang province of three speakers producing open syllables with labial onsets. The pitch excursions of the six tones are shown in Fig. 1. On obstruent-initial syllables, breathy voice is a redundant but still salient characteristic of /bh/-series tones: including H1*-H2* at vowel onset as a predictor in a linear discriminant analysis (LDA) decreases confusion between tones 5 and 6 by around 15% over a function employing only F0-based predictors. While this suggests analyzing voice quality as an integral part of the /bh/-series tones, the 'same' tones, with the same pitch shapes, are emphatically *not* breathy when they occur on sonorant-initial syllables: here, H1*-H2* is significantly lower (β=-4.93, $t(517)=-11.34$, $p<0.0001$) and using it in addition to (or in place of) F0-based predictors *decreases* LDA classification accuracy by 5-10% per tone. On the other hand, there is some evidence that breathiness is related to the phonological status of /bh/. While two speakers show a three-way VOT contrast (prevoiced /b/~/bh/, voiceless unaspirated /p/ and voiceless aspirated /ph/), one speaker consistently produces /bh/ as voiceless unaspirated [p] – but distinct from /p/ in terms of voice quality (β=7.7, $t(112)=12.09$, $p<0.0001$).

In summary, F0, VOT, and breathiness are all potentially active cues to CBT tone categories but are active to different degrees in different portions of the tone system. While the six tones are distinguished solely by pitch (height and contour) in syllables with sonorant onsets, they are signaled by all the three acoustic properties in syllables with initial obstruents. In such a system, breathiness is the phonetic realization of a tonal feature underlyingly linked to initial voiced obstruents but re-associated with tones on the surface (cf. Svantesson, 1983). Moreover, the four-way laryngeal contrast among initial obstruents is cued differently for different speakers. This initial acoustic exploration thus reinforces the fact that tone is not only (or even primarily) about pitch, but that pitch is simply one aspect of a complex of laryngeal features. While further work is clearly needed, CBT, as a language preserving a rare a stage of tonogenetic transition, illustrates the highly dynamic character of phonetic cues in complex tone systems.

References

Haudricourt, A.-G. 1965. Les mutations consonantiques des occlusives initiales en mon-khmer. *Bulletin de la Soc. Ling. de Paris* 60, 160-172. / Hoàng Văn Ma. 1997. The sound system of the Tày language of Cao Bang Province, Vietnam. In: Edmondson, J., Solnit, D. (eds.), *Comparative Kadai: the Tai branch*. Dallas: SIL, 221-231. / Pittayaporn, P. 2009. *The phonology of Proto-Tai*. Cornell University PhD thesis. / Svantesson, J. 1983. *Kammu phonology and morphology*. Lund: CWK Gleerup.

Intra-oral Pressure as Independent of Duration: Stops in Bininj Gun-wok

Hywel Stoakes
The University of Melbourne
hstoakes@unimelb.edu.au

This study reports the results of an investigation into phonologically contrastive medial stops in Bininj Gun-wok (BGW), an Australian Language spoken in Australia's Northern Territory. The results show that there are differences in the integral of peak pressure measured as a function of time in medial stops. The phonologies of Australian languages are remarkably similar despite significant linguistic variation. A single stop series with no phonological voicing contrast is the prototypical configuration amongst the majority of Australian languages, although rare cross-linguistically (Dixon, 2002). There are a number of Australian languages, including BGW, which have a demonstrated double stop series. The medial stops in BGW differ phonetically along a number of dimensions, primarily duration rather than voicing and inter-articulator timing. There are however other phonetic correlates that may be independent of voicing and voice onset time, signaled by differences in the magnitude of peak pressure and pressure impulse. Articulatory strength has been cited as a reason to apply the labels lenis and fortis to these stop categories (Evans, 2003). The descriptions lenis and fortis are not well defined cross-linguistically from a phonetic perspective, although the labels are widely applied in languages in which voice onset time does not sufficiently separate the stop series (Ladefoged & Maddieson, 1996). In Jawoyn, a language related to BGW, Jaeger (1983) has argued that pressure differences in medial stops are not independent of duration. It has been noted previously in other Australian languages that there is a difference between the rise rates of the intra-oral pressure between the categories with fortis stops showing a much more rapid rise in the intra-oral pressure than lenis stops (Butcher, 1992). Malécot (1966, 1970) introduced a measurement that he termed the Pressure Impulse (PI)—the integral of the peak pressure curve measured as a function of time with limits as the onset and offset of oral closure in a plosive consonant. This is reported in the current study using the derived units Pascal seconds (Pa·s). In BGW the polysynthetic nature of the language forms many homorganic clusters that cross morpheme boundaries and these geminated stop clusters are said to contrast with the fortis stops category (Evans, 2003).

In this study, six—three male and three female—first language speakers of the Kunwinjku variety of Bininj Gun-wok were recorded under field conditions in a remote community in Western Arnhem Land, Australia. This study utilizes both acoustic and aerodynamic field recordings of intra-oral pressure (P_o) using standard aerometric techniques (Ladefoged, 2003). All intra-oral pressure measurements use real Kunwinjku words containing medial bilabial stops contained within a carrier phrase. Results of this study show that there is a durational difference between lenis and fortis stops with fortis stops on average 53 ms longer than lenis stops ($p > .001$). Lenis stops have a mean duration of 161 ms and lenis stops have a mean duration of 76 ms. Additional to duration there are differences in the peak pressure timing. Lenis and fortis stop reach their peak pressure at a similar time—at approximately 80 ms after closure—but the peak pressure for fortis stops is just under twice that of lenis stops—450 Pa for fortis vs. 200 Pa for lenis suggesting a difference between the peak pressure targets. A statistical analysis of the pressure impulses of lenis, fortis and geminate stops using linear mixed effects modeling (χ^2 (9, N = 148) = 28, $p < .001$) confirms that there is a significant difference in the pressure impulse between the categories. Testing a main effect of stop category shows a significant when lenis is compared with fortis ($p < .001$) and also a significant difference when fortis is compared with geminates ($p < .001$). The mean difference between the PIs shows fortis stops have mean PI of 49 Pa·s higher than lenis stops (55 Pa·s compared with 6 Pa·s). Fortis stops have a mean PI of 30 Pa·s higher than that of geminate clusters (25 Pa·s).

In medial stops, the lenis category is a rescaled version of fortis stops, each having a different peak pressure target. This is demonstrated by the timing and pressure impulse measurements. There are pressure impulse differences between fortis stops and geminates with the former patterning with heterorganic clusters. This finding suggests that duration is an independent parameter to intra-oral pressure for BGW.

References

Butcher, A. R. 1992. Intraoral pressure as an independent parameter in oral stop contrasts. In Pittam, J., ed., *The Proceedings of the 4th Australian International Conference on Speech Science and Technology.*, pages 286–291, Canberra. Australian Speech Science and Technology Association. / Dixon, R. M. W. 2002. *Australian Languages: Their Nature and Development*. Cambridge CUP. / Evans, N. 2003. *Bininj Gun-Wok: a pan-dialectal grammar of Mayali, Kunwinjku and Kune*. Australian National University, Canberra. / Jaeger, J. J. 1983. The fortis/lenis question: evidence from Zapotec and Jawoñ. *Journal of Phonetics*, 11:177–189. / Ladefoged, P. 2003. *Phonetic Data Analysis: An Introduction to Fieldwork and Instrumental Techniques*. Blackwell, Malden / Ladefoged, P. and Maddieson, I. 1996. Sounds of the World's Languages. Blackwell, Oxford / Malécot, A. 1966. The effectiveness of intra-oral air-pressure-pulse parameters in distinguishing between stop cognates. *Phonetica*, 14: 65–81. / Malécot, A. 1970. The lenis-fortis opposition: its physiological parameters. *The Journal of the Acoustical Society of America*, 47(6B): 1588–1592.

Testing the Syllabicity of Inserted Vowels: A Word-game Experiment in Pnar

Hiram Ring[1]*, James Gruber*[2]*, Jacqui Nokes*[3]

[1]Nanyang Technological Institute, [2]New Zealand Institute of Language Brain and Behaviour, [3] University of Canterbury

hiram1@e.ntu.edu.sg, james.gruber@canterbury.ac.nz,
jacqueline.nokes@canterbury.ac.nz

This study investigates a pattern of vowel insertion that occurs in complex onsets of the Pnar language. We ask what the phonological nature of the inserted vowel is and whether it is lexically stored or if it reflects a property of Pnar onset phonotactics? Evidence comes from a word-game experiment that tests listeners' intuitions regarding the syllabicity of real and pseudo-word forms. Results indicate that listeners did not regard the inserted vowel as syllabic and that their judgments were influenced by properties of the complex onset. The data help us to understand when and how listeners reconstruct phonological vowels from minimal vocalic information, when they filter it out, and how their native phonology shapes these perceptual strategies.

Pnar is a Khasian language of the Mon-Khmer family spoken in the northeast Indian state of Meghalaya. Data for this study come from fieldwork conducted in Jowai, the largest town in eastern Meghalaya and center for a prestige dialect. Pnar (and Khasian languages generally) is notable in its phonotactics for allowing a wide variety of initial onset clusters, including many which represent marked sonority sequences (1-6). Some clusters are regularly broken up a short vocalic element (thus, CᵊC). This process should not be understood as a repair for illicit sonority curves because it systematically occurs within sonority rises (7, 8 below) while it is not employed in more marked sonority plateaus (e.g. [kti] and not [*kᵊti] in (1)).

(1)	**kt**i	'*hand*'		(3)	**k**ᵊ**b**a	'*rice*'		(5)	**d**ᵊ**kʰ**ot	'*member*'	(7)	**b**ᵊ**l**a͡ɪ	'*god*'
(2)	**ps**iaʔ	'*enter*'		(4)	**p**ᵊ**n**a	'*buttocks*'		(6)	**r**ᵊ**kʰ**a͡ɪ	'*laugh*'	(8)	**d**ᵊ**ŋ**ɛm	'*bear*'

Davidson & Stone (2003), Davidson (2005), and Hall (2006) have demonstrated that some epenthetic vowels are unlike their lexical vowel counterparts, in that they have different articulatory sources and subsequent acoustic manifestations. Hall (2006) uses the term "intrusive" to describe vowels which are emitted between non-overlapping consonants as opposed to those produced by a distinct vowel-forming gesture. Acoustic evidence in Ring (2012) suggests the vocalic segment in Pnar is intrusive in this sense – inserted vowels had a shorter duration and different vowel quality from other vowels he measured. A clearer depiction of the phenomenon in Pnar and of the role that epenthetic vowels may play in syllabification can be acquired by understanding speaker intuitions regarding syllabification of these complex clusters.

A word-game experiment, based on the reversal game in Donselaar, Kuijpers and Cutler (1999), tested the processing and syllabification of the intrusive vowel and a comparison set of weak central vowels. Thirty-two native speakers were taught to reverse the order of phonemes in a monosyllabic word, but reverse the syllable order in disyllabic forms (TAM yields [mat], BATAM yields [tamba]). Stimuli were presented auditorily and consisted of experimental tokens and filler tokens. Experimental tokens allowed us to test (a) real and pseudo-words with onset clusters and (b) real and pseudo-words with an initial unstressed lexical vowel in a CVCV(C) structure, and (c) real and pseudo-word variants of words in (a) and (b) with the intrusive vowel spliced in (when not expected) or out (when expected). Presentation of the altered stimuli were balanced across subjects such that no subject heard both the original and spliced version of an individual token.

Results indicate that Pnar words with onset clusters and an intrusive vowel are treated as monosyllables – 87.2% of real and pseudo-words. In comparison, with a weak initial vowel elicited categorically disyllabic responses – 97.1% across real and pseudo-word forms. Overall, real and pseudo-word stimuli were treated similarly, suggesting that the intrusive vowel is a product of certain phonological patterns rather than part of a speaker's lexical knowledge. The digitally-altered stimuli had an uneven effect across the conditions in that (a) splicing a vowel out did not affect responses but (b) splicing an intrusive vowel into a onset cluster which typically did not attest such a vowel created ambiguity.

We conclude that the vocalic segment is not underlying and is phonologically null. Still, it constitutes part of a phonetically well-formed production of Pnar onsets that has not been phonologized. Implications of the study concern the typology of epenthetic vowels and of 'sesquisyllabicity', a prevalent structure in many related and geographically proximate languages (Thomas, 1992).

References

Davidson, L. 2005. Addressing phonological questions with ultrasound. *Clinical Linguistics and Phonetics* 19:6/7, 619-633. / Davidson, L., Stone, M. 2003. Epenthesis versus gestural mistiming in consonant production: an ultrasound study. In *Proc. of the WCCFL* vol. 22. / van Donselaar, W., Kuijpers, C., Cutler, A. 1999. Facilitatory effects of vowel epenthesis on word processing in Dutch. *Journal of Memory and Language* 41, 59-77. / Hall, N. 2006. Cross-linguistic patterns of vowel intrusion. *Phonology* 23, 387-429. / Ring, H. 2012. A phonetic description and phonemic analysis of Jowai-Pnar. *Mon-Khmer Studies* 40, 133-175. / Thomas, D. 1992. On sesquisyllabic structure. *Mon-Khmer Studies Journal* 21, 206-210.

Wh-question Intonation Patterns in the Showamura Dialect of Japanese

Jason Ginsburg, Ian Wilson**, Emiko Kaneko**, Naomi Ogasawara***

*Osaka Kyoiku University, **University of Aizu

jginsbur@gmail.com, wilson@u-aizu.ac.jp, kaneko@u-aizu.ac.jp, naomi-o@u-aizu.ac.jp

This paper presents some characteristics of wh-questions that are used in the dialect of Japanese spoken in Showamura, which is a small rural village in the Aizu region of Fukushima prefecture in Japan. The traditional dialects of the Aizu region of Japan are endangered. Young people of this region are not fluent in these dialects. We collected data via interviews with 26 Showamura residents (average age 77.5), and we compiled a corpus of this data. We present and analyze findings regarding intonation in wh-questions in the Showamura dialect.

According to Ishihara (2003), in wh-questions in standard (Tokyo) Japanese, there is a pitch peak on a wh-phrase followed by post-focus reduction (PFR). For example, when comparing a statement containing the indefinite *nanika* 'something' with a corresponding wh-question containing the wh-phrase *nani-o* 'what-Acc', Ishihara finds a higher pitch peak on the wh-phrase than on the indefinite, and in addition, after the wh-phrase, there is a "PFR effect [that] spreads to all post-FOCUS [post-wh] material" (Ishihara, 2003, p. 52). We found a different pattern in the Showamura dialect, in that a) a pitch peak does not always fall on a wh-phrase, and b) a wh-phrase is not always followed by PFR.

Example (1) is a statement with the indefinite *nanika* 'something' (Fig. 1) and (2) is a corresponding wh-question, from the same speaker, with *nani-o* 'what-Acc' (Fig. 2). Note that there is a pitch peak on both the indefinite *nanika* (Fig. 1) and the wh-phrase *nani-o* (Fig. 2). However, there is no significant PFR effect after the wh-phrase. Rather there is only a slight drop in pitch. There is a greater drop in pitch after the indefinite in Fig. 1 than after the wh-phrase in Fig. 2.

(1) Jisama-ga ***nanika*** *nagashi-de kuteta yo*
grandpa-Nom soething kitchen-Loc ate Emph
Grandpa ate something in the kitchen.

(2) *Jisan-wa* ***nani-o*** *nagashi-de kuta no-gana*
grandpa-Top what-Acc kitchen-Loc ate Q
'What did grandpa eat in the kitchen?'

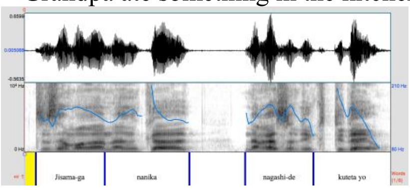


Fig. 1 (80 year-old male)

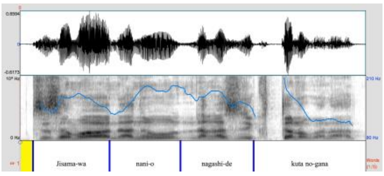


Fig. 2 (80 year-old male)

We found two primary patterns with respect to pitch peaks on wh-phrases - either there is a pitch-peak on the wh-phrase (70 out of 116 examples = about 60%) or there is rising intonation on the wh-phrase (46 out of 116 examples = about 40%). Example (3), as shown in Fig. 3, is a wh-question with a pitch peak on *ga* 'Nom' (the particle following the wh-word). There is a notable lack of a PFR effect, as there is a peak on the next phrase *miru* 'see'. Example (4), from the same speaker, has rising intonation (but no peak) on the wh-phrase *nani* 'what', with the pitch peak falling on the following verb *nondeta* 'drank' (Fig. 4). In (5), there is a small pitch peak on the wh-phrase *donna* 'which'. However, there is almost no drop in pitch until after the noun *kutsu* 'shoes' (Fig. 5).

(3) ***nani-ga*** *miru*
what-Nom see
'What can you see?'

(4) *yube na basama yuruipata-de* ***nani*** nondeta no-ya
evening Emph grandma hearth *Loc* what drank Q
'That evening, what did grandma drink in the hearth?'

(5) *Otoko-no-ko* ***donna****-iro-no-kutsu haiteiru*
boy which-color-Gen-shoes wearing
'Which color shoes is the boy wearing?

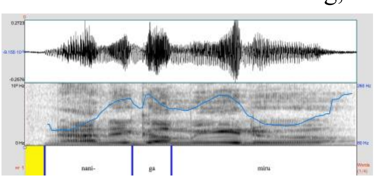


Fig. 3 (75-year old male)

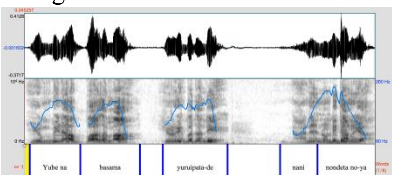


Fig. 4 (75-year old male)

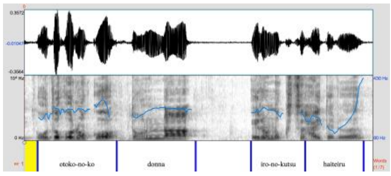


Fig. 5 (75-year old female)

These data suggest that the typical intonation patterns found in Standard Japanese do not necessarily hold in the Showamura dialect. We provide further analyses of this and related data to demonstrate that the prominence of a wh-phrase can be indicated with either a pitch peak or with rising intonation, and in addition, there is no requirement for PFR following a wh-phrase.

References

Ishihara, S. 2003. *Intonation and Interface Conditio*ns. Massachusetts Institute of Technology Dissertation.

A study of the prosodic encoding of TOPICS as a category of information structure in Ngarinyman, a language of Australia

Candide Simard[a], Faith Chiu[b], Eva Schultze-Berndt[c]

School of Oriental and African Studies[a], University College London[b], University of Manchester[c]

cs75@soas.ac.uk[a], faith.chiu.11@ucl.ac.uk[b], eva.schultze-berndt@manchester.ac.uk[c]

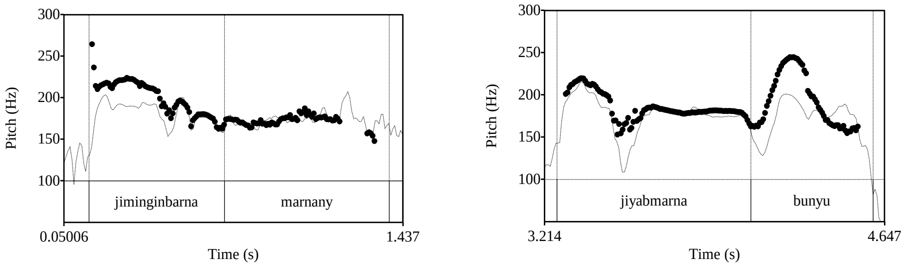
Prosody serves important communicative functions, notably that of marking of Information Structure (IS hereafter) categories such as (1) topic, which refers to what a sentence is about and typically represents given information, and (2) focus, which typically represents new information about the topic (Lambrecht, 1994; Vallduví and Engdahl, 1996). This paper reports on a study of the prosodic encoding of TOPICS in Ngarinyman, a severely endangered language spoken in the Northern Territory of Australia, based on data extracted from a corpus of spontaneous language - or at least non-read speech - collected during fieldwork with the few remaining fluent speakers, making use for their analysis of methods and tools originally confined to the laboratory.

We follow Krifka (2007), Lambrecht (1994) and others in defining Topic in terms of 'aboutness', i.e. as the constituent with a referent about which the sentence provides some information. Our annotation system is based on the methodology developed in the QUIS project (Questionnaire on Information Structure; Skopeteas et al. 2006) and takes into account the information status (degree of accessibility) of the topic as well as its position (left vs. right edge) and its prosodic and syntactic integration into the main clause. We also include in our investigation contrastive topics, considered as a subset of aboutness topics, as shown in example 1, where the contrastive topics are *jimingin =barna* 'kidneys' and *jiyabmarna* 'liver'.

Although the conceptual scaffolding is fairly straightforward, identifying topics in natural discourse is notoriously difficult (Cook and Bildhauer 2011). We will show that in our data, cues taken from prosody, segmental marking, constituent order as well as discourse context all contribute to the identification of topics. For example, Ngarinyman has grammatical markers that only appear on some discourse-given topics and can thus be considered a sufficient but not necessary criterion for the identification of a constituent as topical.

Figure 3. Contrastive topics: the thick line shows the pitch, the dotted line is the intensity

(1) jimin.gin=barna marnany [jiyab=ma=rna]TOP [bunyu]COM
kidney=1SG.S bad liver=GIVEN=1SG.S good
'my KIDNEYS are sore; my liver is all right' (lit.'kidneys I am sore; liver I am good') ER,ES08_A06_02.157-8)



The aim of this research is to show the prosodic characteristics of topical constituents, and to further investigate whether subtypes of topics, semantically defined, are distinguished by prosodic means.

An instrumental analysis couched within the framework of PENTA model (Xu, 2005) is conducted on rigorously selected datasets for each sub-type of TOPICS. The results of the measurements of the correlates of pitch (F0), intensity and duration are further tested with PENTAtrainer2 (Prom-on and Xu, 2012), a modelling tool used to validate our hypotheses. We find that the categories of given TOPICS differ significantly from that of contrastive topics which bear marking similar to that of focus. Finally, the comparison in the measurements of the subcategories of given TOPICS suggests that these may be gradual rather than truly categorical distinctions.

References

Cook, P., Bildhauer, F. 2011. Annotating Information Structure: The case of topic. In Dipper, S., Zinsmeister, H. (eds.), *Beyond Semantics. Corpus-based Investigations of Pragmatic and Discourse Phenomena*, 45-56. Bochum: Universität Bochum: Bochumer Linguistische Arbeitsberichte 3. / Lambrecht, K. 1994. *Information structure and sentence form. Topic, focus and the mental representations of discourse referents*. vol. 71: Cambridge studies in linguistics. Cambridge: Cambridge University Press. / Krifka, M. 2007. Basic notions of information structure. *Interdisciplinary Studies of Information Structure* 6: 13-56. / Prom-on, S., Xu, Y. 2012. PENTATrainer2: A hypothesis-driven prosody modeling tool. In Proceedings of the Fifth ISEL Conference on Experimental Linguistics (ExLing 2012), Athens, Greece. Pp. 93-100. / Skopeteas, S., Fiedler, I., Hellmuth, S., Schwarz, A., Stoel, R., Fanselow, G., Fery, C., Krifka, M. (eds.) 2006. *Questionaire on information structure (QUIS).* Potsdam: University of Potsdam. / Vallduví, E., Engdahl, E. 1996. The Linguistic Realization of Information Packaging. *Linguistics* 34.459-519. / Xu, Y. 2005. Speech melody as articulatorily implemented communicative functions. *Speech Communication* 46: 220-251.

ERP correlates of two types of implicit knowledge of probabilistic phonotactics

Claire Moore-Cantwell and Lisa Sanders
University of Massachusetts Amherst
cmooreca@linguist.umass.edu, lsanders@psych.umass.edu

Linguistic knowledge consists of a lexicon of memorized forms as well as a system of generalizations used in the production of novel outputs. These generalizations can be formalized as abstract grammatical rules or constraints, or as an epiphenomenon of the organization of the lexicon (e.g. into networks, Bybee, 1995, Pierrehumbert, 2000). Dual-route (Pinker, 1999) and two-system (Ullman, 2004) models divide phonological knowledge up into categorical generalizations formalized as abstract rules, and probabilistic generalizations which are the result of analogy over lexical items. We use behavioral and neurological evidence to argue instead that speakers have abstract knowledge of some probabilistic generalizations and 'epiphenomenal' or lexicon-based knowledge of others.

It is well-known that the trend for word-initial stress in English plays a role in lexical processing (e.g. Cutler (1987)). We tested two more complicated trends. In English words longer than two syllables, stress is typically penultimate ('banána') or antepenultimate ('Cánada'). A search of the CMU pronouncing dictionary (Weide, 1994) revealed that the rate of occurrence of each stress position depends on the quality of the word-final vowel, with [i]-final words biased towards taking antepenultimate stress, and [ə]-final words biased towards penultimate stress. In three-syllable, monomorphemic words, all of whose syllables are light, [i]-final words take antepenultimate stress about 80% of the time, and [ə]-final words take penultimate stress about 70% of the time. When the search space is expanded to include all words three syllables long or longer, both monomorphemic and morphologically complex, the [i]-final trend becomes more robust, with the trend-observing pattern occurring 88% of the time, while the [ə]-final trend becomes less robust with the trend-observing pattern occurring only 54% of the time.

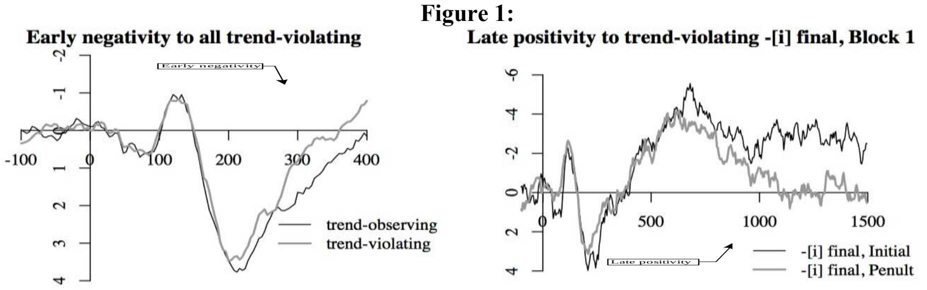
Final V	Trend-observing	Trend-violating
[I] (more robust)	**(A) Antepenult.:** récipe spághetti bámaki	**(C) Penultimate:** spaghétti recípe bamáki
[ə] (less robust)	**(B) Penultimate:** vanílla nebúla famáka	**(D) Antepenult.:** nébula vánilla fámaka

Items: 40 English words ('spaghétti'), their mis-stressed counterparts ('spághetti'), and 80 nonce words ('bámaki') were presented auditorily. There were equal numbers of each cell (**A,B,C,D**) and all items were three syllables long.

Procedure: 21 participants listened to items and subsequently rated them on a 1-4 scale based on how likely they were to be an actual word of English. 1280 items were divided into 5 blocks. Meanwhile, EEG was recorded continuously using a HydroCel Geodesic Sensor Net, and segmented into 1500ms epochs timelocked to the onset of each stimulus. Statistical analyses were performed on mean amplitudes within the time windows of interest.

Results: [i]-final penultimate nonwords were rated as less word-like than all other shapes ($F(2,42)=4.44$, $p<0.05$). Both types of [ə]-final nonwords were rated as equally word-like. Very early after word onset (280-380ms) a greater negativity was found for trend-violating (**A**, **B**) than for trend-observing (**C**, **D**) words and mis-stressed words ($t(20)=-2.31$, $p=0.02$). In mis-stressed words only, this negativity begins even earlier, on the N1 auditory onset peak (120-180ms), where its size is modulated by acoustic differences due to stress position ($t(20)=4.156$, $p<.05$). In the first block of the experiment, [i]-final penultimate (**C**) nonwords elicited a late positivity (1000-1500ms) relative to their antepenultimately stressed counterparts (**A**) ($t(20)=-2.73$, $p=<.05$), similar in timing and morphology to a late positive component typically elicited by categorical phonotactic violations, and to a P600 elicited by syntactic violations. Additionally, [ə]-final antepenultimate (**D**) words and mis-stressed words elicited a larger N400 than their penultimately stressed counterparts (**B**) ($t(20)=4.04$, $p< 0.05$). Surprisingly, there was no corresponding N400 difference between trend-observing and trend-violating [i]-final words. This difference between trends requires further investigation.

Figure 1:



Both the trends we tested affect very early stages of lexical processing, trend-violating words being more difficult to process than trend-observing words. **Only the trend which is more robustly supported by the lexicon is respected in speakers' ratings of nonwords, and is reflected in a late positivity closely related to that elicited by phonotactic violations (Domahs et al., 2008; Pitkanen, 2010), and syntactic violations (e.g. Osterhout and Holcomb, 1992).**

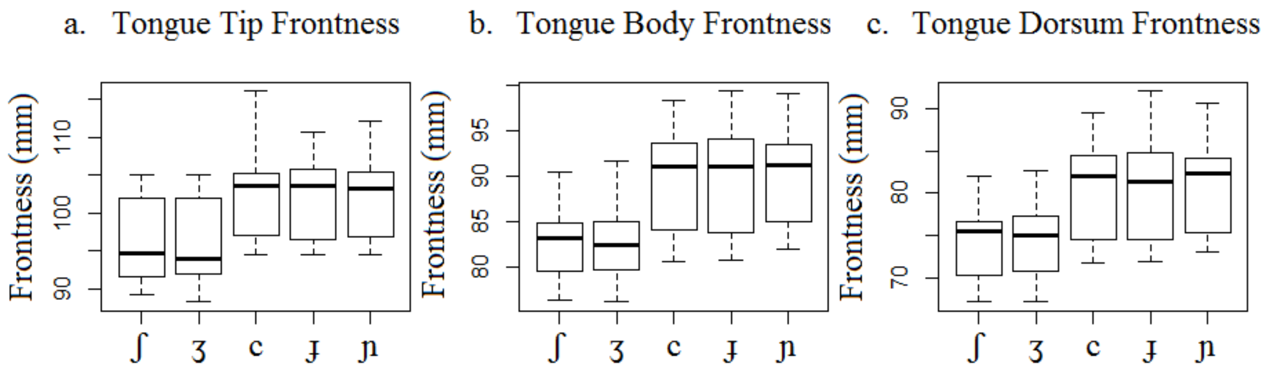
An EMA Examination of Czech Post-Alveolar and Palatal Consonants

Phil Howson & Alexei Kochetov
University of Toronto
phil.howson@mail.utoronto.ca, al.kochetov@utoronto.ca

Consonant articulations produced behind the alveolar ridge have been traditionally divided into two classes palatoalveolars (sibilants /ʃ, ʒ, ʧ, ʤ/) and palatals (/c, ɟ, ɲ, ʎ, ç, ʝ, j/). However, cross-language studies of so-called 'palatals' (e.g. Keating, 1988; Recasens, 1990; 2013) suggested that these include at least two distinct places - alveolopalatals and palatals proper - and thus should be classified accordingly. For the alveolopalatals, the closure spans from the alveolar region into the palatal region, involving a constriction made with the tongue blade and dorsum. This constriction is expected to be more extensive and possibly more anterior than for palatoalveolar fricatives (ʃ, ʒ) articulated with the tip or blade in the post-alveolar region. Czech presents an interesting case study, as it has a set of contrastive palatoalveolars and palatals in its phonetic inventory, including /ʃ, ʒ, c, ɟ, ɲ/. The goal of this paper is to determine the relative position of the tongue in the articulation of these consonants, testing Recasens' proposal concerning the distinct alveolopalatal place of articulation.

Electromagnetic articulography (EMA) data with Czech phonemes /ʃ, ʒ, c, ɟ, ɲ/ were collected from 6 native speakers of the language. 3 sensors were attached to the participants' tongue along the mid-sagittal plane – on the tongue tip, the tongue body and the tongue dorsum. 15 tokens of each phoneme were produced in nonce words in three distinct positions: word initial (#_a), intervocalic (a_a), and word final (a_#). Measurements were taken at the point of maximum constriction using Mview (Tiede, 2013), a toolbox for MATLAB. A repeated measures ANOVA was performed with the factors Consonant (5 levels: ʃ, ʒ, c, ɟ, ɲ) and Environment (3 levels: #_a, a_a, a_#). The results indicated a significant difference between /ʃ, ʒ/ and /c, ɟ, ɲ/ along the mid-sagittal plane (Figure 1). Specifically, the entire tongue is further forward in the mouth for the so-called palatals than for the palatoalveolars. Furthermore, the height of the tongue body and tongue dorsum were higher for former than for the latter. The environment did not significantly affect the consonants' articulation.

Figure 1. a-c. Frontness of the tongue articulator, higher numbers indicate a more anterior articulation.



The results suggest that the hypothesis put forward by Recasens (2013) - that the consonants traditionally referred to as palatal, /c, ɟ, ɲ/, are in fact alveolopalatal - is correct for the Czech data. The frontness and height of the tongue for /c, ɟ, ɲ/ suggests that the closure extends from the alveolar region into the prepalatal region. This is consistent with Hála's (1923) palatograms for Czech consonants, as well Keating's (1988) general observation that the tongue blade is an active articulator for the alveolopalatals. Keating (1988) also suggests the tongue dorsum is an active articulator, the findings suggest that the tongue body and dorsum are active articulators. The tongue body is raised to press against the post-alveolar and palatal region to extend the area of constriction; the tongue dorsum is raised even higher than tongue body suggests it is laterally pressing against the palatal and post-palatal regions. These findings are consistent with Hála's (1923) palatograms, which suggest the primary constriction is in the alveolar and post-alveolar region, with substantial contact extending to the molars. Furthermore, the non-significant difference between the consonants /c, ɟ, ɲ/ provides evidence for symmetry requirement hypothesis (Recasens, 2013), as the consonants agree in constriction location although they differ in manner. Finally, the raising of the tongue body during the articulation of the palatoalveolars suggests the tongue body is also involved; the height and the frontness of the tongue tip and body, also suggests that they have a more localized area of constriction, which is assumed to be the post-alveolar region.

References

Hála, B. 1923. *K popisu pražskévý slovnosti: studie z experimentální fonetiky*. V Praze: Nákl / Keating, P. 1988. Palatals as complex segments: X-ray evidence. *UCLA WPP* 69, 77-91 / Recasens, D. 1990. The articulatory characteristics of palatal consonants. *J.Phon* 18, 267-280 / Recasens, D. 2013. On the articulatory classification of (alveolo)palatal consonats. *JIPA* 43(1), 1-22 / Tiede, Mark. 2013. Mview, articulatory visualization software. Haskins Laboratories.

Inductive learning of long-distance dissimilation as a problem for phonology

Kevin McMullin and Gunnar Ólafur Hansson

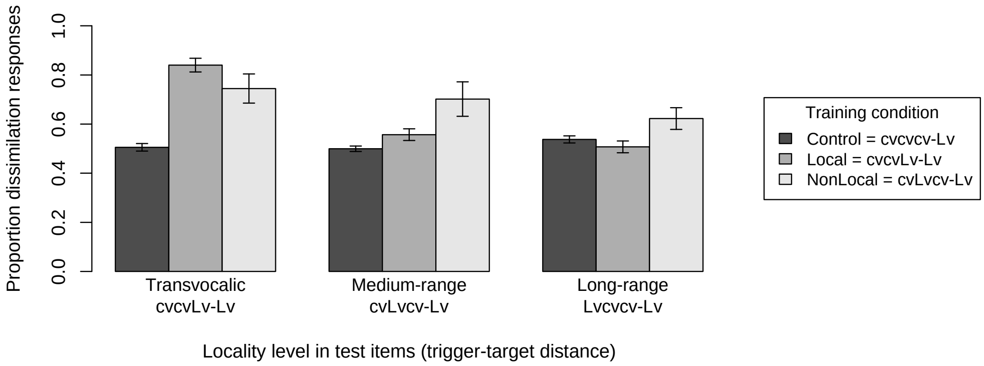
University of British Columbia

kevinmcm@alumni.ubc.ca, gunnar.hansson@ubc.ca

This study investigates the hypothesis that phonological typology is shaped, at least in part, by human learning biases. The typology of consonant harmony, a phonotactic dependency in which two non-adjacent consonants must match for some feature value, includes just two types of languages with respect to the distance between interacting segments. Some languages apply harmony only in *transvocalic* contexts, others are *unbounded* systems that apply harmony over any distance within the harmony domain (Hansson 2010, Rose & Walker 2004). Recent artificial language learning experiments show that this dichotomy is mirrored by human learning biases. Adult participants exposed to harmony in CVC contexts do not generalize to greater distances, but exposure to CVCVC items with harmony results in participants learning an unbounded pattern that generalizes to all distances (Finley 2011, 2012). We report on the findings of an analogous experiment testing whether humans show the same biases when learning long-distance dissimilation.

Participants were given the task of conjugating CVCVCV 'verb' roots from a controlled artificial language into 'past' and 'future' tense by adding the suffixes *-li* and *-ru*, respectively. The language contained a suffix-triggered liquid dissimilation pattern: a liquid [l] or [r] in the root alternated so as to differ from the suffix liquid (*dupire* ~ *dupire-li*, *dupile-ru*). Three groups of 12 participants were trained on different subsets of words adhering to the pattern. In the training data for the Local group, liquids occurred only in the root-final syllable (transvocalic dissimilation), whereas the Nonlocal group encountered roots with liquids only in the second syllable (dissimilation across two syllables). The Control group was not exposed to any liquid roots. All participants were tested on the same set of novel items to determine if they learned a pattern compatible with their training and whether they generalized dissimilation to other positions. Responses were analyzed using a mixed-effects logistic regression model. Results show that the Local group learned a transvocalic pattern that does not apply at greater distances, but that the Nonlocal group learned a genuinely unbounded pattern, generalizing both to lesser and greater distances (Figure 1).

Figure 1: Proportion of dissimilation responses by distance and training condition. Error bars represent standard error based on within-subject means for each category.



This result gives rise to a paradox for current theoretical models of non-adjacent phonotactics. The fact that humans appear to learn harmony and dissimilation relations in the same way suggests a unified phonological analysis for the two. Indeed, recent analyses seek to integrate them (Bennett 2013) using the surface correspondence notion originally proposed for consonant harmony (Rose & Walker 2004). However, this generates a predicted typology of dissimilation that is quite different from harmony. With respect to distance, there should be languages that dissimilate consonants only when more than one vowel intervenes, and no languages should limit dissimilation to transvocalic contexts. If the typologies of harmony and dissimilation are fundamentally different, as predicted by the unified correspondence analysis of the two, then we would expect them to be learned in different ways; our experimental findings contradict this. We consider and contrast two solutions. The first is a modification to the constraints used in correspondence analyses of harmony and dissimilation; the second is a proposal for a computational learning algorithm that operates over tiers (Heinz et al. 2011) and has the same inductive learning biases that humans appear to exhibit in the laboratory.

References

Bennett, W. 2013. Dissimilation, consonant harmony, and surface correspondence. Rutgers Univ. diss. / Finley, S. 2011. The privileged status of locality in consonant harmony. *J Mem Lang* 65:74-83. / Finley, S. 2012. Testing the limits of long-distance learning: learning beyond a three-segment window. *Cog Sci* 36:740-56 / Hansson, G. 2010. *Consonant Harmony: Long-distance Interaction in Phonology*. Berkeley: U. of Cal. Press. / Heinz, J., C. Rawal, and H. Tanner. 2011. Tier-based strictly local constraints for phonology. *Proc. of ACL49*: 58-64. / Rose, S. and R. Walker. 2004. A typology of consonant agreement as correspondence. *Lg* 80:475-531.

~~Asymmetries in cross-linguistic stop/fricative perception~~ Cancelled

Youngja Nam & Linda Polka

McGill University

young.nam@mail.mcgill.ca & linda.polka@mcgill.ca

Previous speech perception research has primarily focused on discriminability between/among different contrasts. Relatively recently, a new line of research has emerged concerning on perceptual asymmetries as a function of stimulus presentation direction within a contrast. With respect to vowels, the natural referent vowel (NRV) framework developed by Polka and Bohn (2011) proposes that directional asymmetries reflect perceptual biases. The NRV framework hypothesizes that a shift from a low to a high salience vowel is easier to detect compared to the reverse direction. For vowels, salience differences are related to position in the vowel space; vowels have more focal spectral energy when adjacent formants become closer which naturally occurs as you move from the center to the corners of the vowel space. To date, directional asymmetries in consonant perception have been overwhelmingly concerned with place of articulation. Only a small number of studies have explored directional asymmetries in consonant manner perception. A few studies show that toddlers perceive a fricative to stop change better than a stop to fricative change (e.g., Altvater-Mackensen & Fikkert, 2010). An adult study reported similar results using an English /ba/-/va/ contrast. Tsushima et al. (2003) found that Japanese listeners discriminated a /v→b/ change better than a /v→b/ change for the non-phonemic contrast. To our knowledge, no published cross-linguistic data are available on directional asymmetries in consonant manner perception. As an extension of the NRV framework to consonants, we hypothesize the existence of acoustic/perceptual salience hierarchy of consonant manners. Specifically, we hypothesize that stops having an abrupt burst and a sharp rise time are acoustically/perceptually more salient than fricatives having long frication and a slow rise time. In this study, we addressed whether directional asymmetries emerge for stop/fricative contrasts other than /b/-/v/ and whether the salience differences are modulated by language experience.

We presented six English stop/fricative contrasts (/tas/-/sas/, /pas/-/fas/, /bas/-/vas/, /das/-/zas/, /das/-/ðas/ and /tas/-/θas/) and two Persian stop/fricative contrasts (/kas/-/xas/ and /gas/-/ɣas/) to English, French and Korean L1 listeners in an AX discrimination task. This stimulus set includes phonemic and non-phonemic contrasts for each language group. For each contrast, the AX task included an equal number of different pairs in both directions (stop-fricative & fricative-stop) and same pairs for each phone (stop-stop & fricative-fricative). On each trial listeners heard a pair of syllables and had to decide whether the syllables started with the same consonants or two different consonants.

With respect to different pairs, no direction differences emerged for /tas/-/sas/ which is native in all 3 language groups. For non-phonemic contrasts, as predicted, listeners in all groups showed better discrimination on fricative-stop pairs than stop-fricative pairs. For non-phonemic contrasts, performance asymmetries were also found for same pairs; listeners were more accurate on stop-stop pairs than fricative-fricative pairs. Further, these patterns emerged for certain phonemic contrasts, e.g., English listeners' perception of /das/-/ðas/ and /tas/-/ θas/.

Taken together, these results concur with our salience hierarchy hypothesis that stops with abrupt onsets are perceptually more salient than fricatives with gradient onsets and thus, listeners' discrimination of stop-fricative pairs is depressed relative to their discrimination of fricative-stop pairs due to salience overriding. Performance asymmetries on same pairs also shows that stop perception is more stable than fricative perception, lending further support to our hypothesis of the acoustically/perceptually privileged status of stops over fricatives. That asymmetries emerged for some native phonemic contrasts suggests that inherent salience differences may not be completely modulated by language experience. Directional asymmetries in adult consonant perception have been typically explained by the featurally underspecified lexicon (FUL) model. For example, Cornell et al. (2013) interpreted their mismatch negativity results as suggesting that when the nasal /n/ serves as the rare deviant amidst the stop /d/ standards, no conflict occurs between the underspecified lexical representations of /d/ and the surface representations of /n/. In the reversed condition, the specified status of the frequent standard nasal stimuli induces a conflict with the deviant stop's surface acoustic representations. However, our results appear to be difficult to be explained by the FUL model. In fact, the FUL model seems to predict higher sensitivity to the fricative-fricative pairs than the stop-stop pairs because as long as the underspecified manner serves as the standard, perceptibility of the deviant stimulus becomes reduced. Overall, our results dovetail conceptually with the NRV framework, suggesting that this framework is also applicable to consonants.

References

Cornell, S. A., Lahiri, A., & Eulitz, C. (2013). Inequality across consonantal contrasts in speech perception: Evidence from mismatch negativity. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(3), 757.

Polka, L., & Bohn, O. S. (2011). Natural Referent Vowel (NRV) framework: An emerging view of early phonetic development. *Journal of Phonetics*, *39*(4), 467-478.

Tsushima, T., Shiraki, S., Yoshida, K., & Sasaki, M. (2003). On stimulus order effects in discrimination of nonnative consonant contrasts. *Acoustical Science and Technology*, *24*(6), 410-412.

Altvater-Mackensen, N., & Fikkert, P. (2010). The acquisition of the stop-fricative contrast in perception and production. *Lingua, 120*(8), 1898-1909.

Acoustic Disjuncture in Consonant Clusters and Vowel Epenthesis

Suyeon Yun
Massachusetts Institute of Technology
suyeon@mit.edu

In loan adaptation, vowel epenthesis frequently occurs as a repair, when a cluster of a source language is phonotactically illegal in the borrowing language. A notable earlier finding has been that the position of epenthetic vowels differs depending on the sonority profile of cluster; clusters with rising sonority, especially stop-sonorant, are more likely to be split by an epenthetic vowel than clusters with falling sonority, especially sibilant-stop (Broselow 1992, Fleischhacker 2001, 2005, Gouskova 2003, Steriade 2006), e.g., 'plastic' → [b**i**lastik] vs. 'study' → [**i**stadi] (Egyptian Arabic). Based on typological and experimental evidence, this paper argues that it is not the sonority profile but acoustic properties of the cluster that determine the epenthesis site. A cross-linguistic survey of 51 languages was conducted. One result is that clusters with level sonority do not behave uniformly, suggesting that other factors than the sonority profile of the cluster determine the position of epenthesis; stop-stop clusters always undergo internal epenthesis (e.g., tkan' (Russian) → [t**i**kan] 'fabric' (Ewen; Choi 2010)), whereas nasal-nasal clusters may undergo external epenthesis (e.g., mnemonicheskij (Russian) → [**y**mnemonicheskij] 'mnemonic' (Kirghiz; Gouskova 2003) vs. mnemosine (Greek) → [m**ɨ**nemosine] 'mnemosyne' (Korean). Second, depending on the acoustic properties of the clusters in the source language, the pattern of epenthesis differs. Korean, for example, borrows English word-final stop-stop clusters, where the first stop is optionally released, with single external epenthesis (e.g., 'compact' → [k^{h}əmp^{h}ekt^h**ɨ**]), but it borrows Russian stop-stop clusters, where both of the stops are obligatorily released, with double epenthesis (e.g., relikt → [lellik^h**ɨ**t^h**ɨ**]). The generalization from this survey is that an epenthetic vowel goes after a stop, or a nasal preceding a nasal or liquid (e.g., #tk→#tək, #kl→#kəl, #mn→#mən (Korean), #ml→#məl), and before a nasal preceding an obstruent or nasal, or a liquid (e.g., #mb→əmb, #mn→əmn (Kirghiz), #lb→əlb). Clusters with a word-initial fricative show complex patterns that are incorporated into the full analysis but not discussed here for simplicity.

The clusters where internal epenthesis is attested involve some acoustic discontinuity between the consonants, but the clusters where external epenthesis is attested do not. So I argue that the presence or absence of acoustic disjuncture (AD) plays a key role in determining the site of epenthesis. I define AD as perceptually salient acoustic discontinuity between the two consonants, one that involves abrupt spectral change accompanied by a rise of intensity. ADs are triggered by the release of stops and nasals. If the input cluster contains an AD, epenthesis occurs at the AD site, e.g., after the released stop or nasal. If there is no AD, epenthesis occurs outside the cluster, e.g., before the unreleased nasal or liquid, or no epenthesis occurs, e.g., after the unreleased stop. I hypothesize that this is because the epenthetic vowel replacing the AD is a perceptually less salient change from the original cluster than epenthesis in a non-AD site.

To test the hypothesis, one production and two perception experiments were conducted. First, nonce words including 31 consonant clusters and their counterparts with epenthetic vowels were recorded from 6 speakers of Russian, a language known for its rich initial and final clusters. The recording notably showed inter- and intra-stimulus variation in the nasal-initial clusters; nasal-obstruent clusters ([mk], [mf], and [ms]) rarely had audible release, i.e., no AD, while nasal-liquid clusters ([ml] and [mr]) always involved audible vocalic release, i.e., AD. For nasal-nasal clusters ([mn]), some were audibly released, i.e., AD, while others were not, i.e., no AD. There was little variation in other clusters. The stimuli for the perception experiments were chosen from these recordings, including released and unreleased stops and nasals in the cluster. To see the effect of the absence of stop release, there were items with word-final stop-C clusters where the release burst of the stop was removed. 16 English speaker subjects participated in the ABX discrimination task, in which each target item containing a cluster was paired with a corresponding item with an epenthetic vowel (e.g., [kmat]-[k**ə**mat]). With the same stimuli, a transcription task was also run, in which 13 English speakers heard the stimuli and transcribed them using English orthography. Results show that the discrimination was worse in the AD trials ($d'=2.7$) than in the non-AD trials ($d'=3.4$), and the effect of AD was statistically significant ($p<.001$), confirming the hypothesis that distinguishing epenthetic forms from clusters involving an AD is more difficult than distinguishing from clusters involving no AD. The results from the transcription task also show patterns similar to the typology. Errors included epenthesis 98% after a stop, 0% before a stop; 84% before a liquid [l], 8% after a liquid [l]; 96% after a released nasal, 0% before a released nasal; 38% after an unreleased nasal, and 47% before an unreleased nasal. These results support the hypothesis that the epenthesis site asymmetry between clusters results from the perceptual similarity between the input cluster and the epenthesized output, depending on the presence or absence of AD.

Selected References

Broselow, E. 1992. Transfer and universals in second language epenthesis. / Fleischhacker, H. 2001. Cluster-dependent epenthesis asymmetries. / Fleischhacker, H. 2005. Similarity in phonology: Evidence from reduplication and loan adaptation. PhD dissertation. UCLA. / Gouskova, M. 2003. Falling sonority onsets, loanwords, and syllable contact. *Proc. CLS* 37: 175-185. / Steriade, D. 2006. Contour correspondence: Evidence from cluster interruption. Handout from OCP 3.

L2 experience can hinder perception of non-native sounds

Jeffrey J. Holliday
Indiana University
jjhollid@indiana.edu

It is typically assumed that as a novice second language (L2) learner becomes more experienced with the L2, the ability to discriminate between L2 sounds should improve or, at the very least, stay the same. It is also typically assumed that even a novice L2 learner should discriminate between L2 sounds at least as accurately as a naïve non-native listener from the same first language (L1) background. These assumptions are supported by studies showing that L2 learners with more experience in the L2 do at least as well on L2 perception tasks as less experienced learners (e.g. Flege & Liu, 2001), and also agree with our basic intuition that L2 learning proceeds linearly from naïve to novice to advanced, with continuous improvement along the way. However, L2 experience comprises multiple types of input, and there is a need to refine our understanding of which parts of L2 experience may promote or hinder L2 learning.

In this paper, we present data suggesting that L2 learners sometimes ignore relevant phonetic detail that naïve listeners from the same L1 background would attend to. Korean has two sibilant fricatives: /s^h/ is heavily aspirated before /a/ but less aspirated before /i/ or /u/, whereas /s*/ is classified as tense and is generally unaspirated in all vowel contexts. Previous research (Holliday, 2012) has shown that naïve L1 Mandarin listeners perceive /s^h/ and /s*/ as members of different categories before /a/, but not before /i/ or /u/. This finding predicts that naïve L1 Mandarin listeners should accurately discriminate between /s^ha/-/s*a/ but not /s^hi/-/s*i/ or /s^hu/-/s*u/. By extension, L1 Mandarin learners of Korean as an L2 – given their experience with Korean – should perform at least as well.

The data reported here are from 50 L1 Mandarin speakers (15 novice and 17 advanced learners of Korean and 18 naïve listeners) and 15 L1 Korean speakers. The novice learners had been studying Korean for less than 10 weeks in Korea, and the advanced learners had studied Korean for at least a year and were enrolled in regular classes at a university in Korea. First, in a perceptual assimilation test, the L1 Mandarin listeners heard CV stimuli with vowel context balanced across /a/, /i/, and /u/. Listeners were asked which Mandarin consonant the consonant portion of each stimulus sounded most similar to in 24 trials of /s^h/, 24 trials of /s*/, and 104 fillers. Second, in a discrimination test, all listeners heard 36 /s^h/-/s*/ trials combined with 72 fillers for a total of 108 trials.

All three L1 Mandarin listener groups categorized both Korean /s^h/ and /s*/ as a sibilant fricative in all vowel contexts except the naïve listeners, who categorized /s^h/ as an aspirated affricate before /a/ in 62.5% of trials. The L1 Mandarin listeners' responses to the /s^h/-/s*/ discrimination trials were analyzed in an ANOVA with a between-subjects factor of listener group and a within-subject factor of vowel context. There was a main effect of vowel [$F(2,92) = 17.47$, $p < .001$] and a significant interaction of vowel and listener group [$F(4,92) = 2.53$, $p = .046$]. Mean discrimination accuracy in the /a/ context was 75.0% for naïve, 58.9% for novice, and 62.7% for advanced listeners, suggesting a drop in accuracy once L2 learning begins. The differences between the naïve listeners and both the novice [$t(30) = 2.96$, $p = .006$] and advanced [$t(32) = 2.18$, $p = .037$] L2 learners were significant. Accuracy levels in the other vowel contexts revealed no such differences (65.7%, 63.9%, and 65.7% for /i/ and 51.0%, 52.8%, and 52.9% for /u/, for naïve, novice, and advanced listeners, respectively).

While it has been suggested that some parts of L2 experience, such as word learning (e.g. Bundgaard-Nielsen et al., 2011), may help learners form new categories, there is also evidence that other parts of L2 experience, such as exposure to orthography, can interfere with L2 learning (e.g. Escudero et al., 2014). Although the exact source of the confusion revealed in this study remains for future investigation, post-test questionnaires revealed that most of the L2 learners believed that Korean /s^h/ is a "light s" and /s*/ is a "heavy s". This belief could stem from Korean orthography, which treats /s^h/ and /s*/ as two types of the same sound (/s*/ is written as a geminate of /s^h/), or from Romanized Korean, which writes /s^h/ as "s" and /s*/ as "ss". Regardless, these results show that fundamental changes take place at the onset of L2 learning, and that these changes do not always result in improved performance in the L2.

References

Bundgaard-Nielsen, R. L., Best, C. T., Tyler, M. D. 2011. Vocabulary size matters: The assimilation of second-language Australian English vowels to first-language Japanese vowel categories. *Applied Psycholing.* 32, 51-67. / Escudero, P., Simon, E., Mulak, K. E. 2014. Learning words in a new language: Orthography doesn't always help. *Bilingualism: Lang. and Cog.* January, 1-12. / Flege, J. E., Liu, S. 2001. The effect of experience on adults' acquisition of a second language. *Studies in Second Lang. Acq.* 23, 527-552. / Holliday, J. J. 2012. *The emergence of L2 phonological contrast in perception: The case of Korean sibilant fricatives*. Ph.D. Thesis, The Ohio State University.

Preferences in Permitted Sequences: A Weighted Markedness Constraint Model

Jeremy Perkins

University of Aizu

jperkins@u-aizu.ac.jp

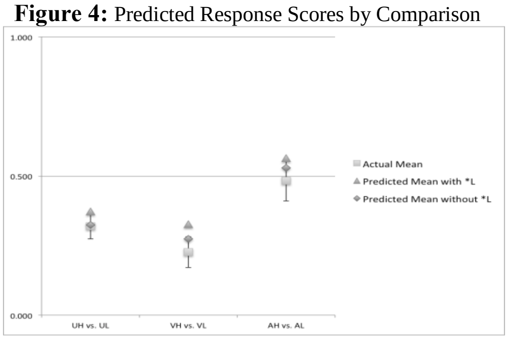
Synopsis: In this talk, I propose a weighted markedness constraint (WMC) model to account for preferences between two grammatical patterns in Thai. Thai speakers show preferences between stimuli in pairs where both have permitted consonant-tone sequences; they prefer voiced-low (VL) and glottal (unaspirated)-low (UL) to aspirated-low (AL) sequences (Perkins 2013). While standard models of grammar account for differences between unattested and permitted patterns, they cannot account for preferences between two permitted patterns. The proposed WMC model explains the pattern of hidden preferences in Thai. In addition, it is shown that the model improves if the constraint *L is removed, supporting previous claims that *L not be included in CON.

Method: In Thai CV: syllables, unaspirated stops and voiced stops are unattested when V: has a high *[dáː] or rising tone *[dǎː]. Native-Thai speaking participants were asked to choose which of two pre-recorded monosyllabic CV: nonce stimuli sounded more Thai-like. Nonce stimuli had one of three tones (high, rising and low) and one of three onsets (aspirated, unaspirated, and voiced stops). Twelve types of comparison pairs were used; these consisted of minimal pairs that differed only in their tone, or in their onset manner. These results were then compared to a predictive WMC model (similar to Coetzee 2004).

Weighted markedness constraint model: In the WMC model, a constraint ranking for a categorical grammar is determined using Biased Constraint Demotion (BCD) (Prince & Tesar 2004), a recursive a learning algorithm. Each stratum is assigned a whole number weighting, ranging from 1 at the bottom stratum, and increasing at intervals of one for each higher stratum. In this study, BCD produced eight constraint strata as an output from 13 markedness constraints adapted from Lee (2011) that analyzed consonant-tone interaction in Thai. For each stimulus pair in the experiment, a predicted response ranging from 0 to 1 is then calculated by summing up the individual weighted contributions of each constraint. This yielded a predicted response, which was then compared with the actual response data. In this model, faithfulness constraints do not play any role because nonce words have no input representation.

Results & Discussion: The predicted responses from the WMC model matched the preferences where Thai speakers preferred VL and UL sequences to AL sequences. Additionally, a large improvement in the fit of the model was observed when the constraint *L was removed. The exclusion of *L is supported by cross-linguistic evidence that L tone is generally less marked than H tone (Yip 2002:41). The result suggests that CON does not contain a markedness constraint, *L. This claim is in accordance with Gouskova (2003), where a markedness constraint for the least marked value is posited to be absent from CON. The results for the UH-UL, VH-VL and AH-AL comparisons with and without *L are overlaid with the results from the experiment, in Figure 1.

Figure 4: Predicted Response Scores by Comparison



Conclusion: In a forced choice task focusing on consonant-tone interaction, Thai speakers' preferences were successfully modeled via a weighted constraint model of markedness constraints. The relative success of a model based on a categorical grammar suggests that markedness constraints are active even in languages where they may play no crucial role in separating the grammatical from the ungrammatical. The results also suggested that the constraint *L does not exist in CON, as its exclusion produced a better fit between the predicted and actual responses.

References

Coetzee, A. 2004. *What it means to be a loser: Non-optimal candidates in Optimality Theory*. Ph.D., UMass-Amherst. Gouskova, M. 2003. *Deriving Economy: Syncope in Optimality Theory.* Ph.D. UMass-Amherst. Lee, S. 2011. Tonal OCP and consonant-tone interaction in Thai. *Journal of SEALS* 4, 61–76. Perkins, J. 2013. *Consonant-tone interaction in Thai*. Ph.D dissertation, Rutgers University. Prince, A. and Tesar, B. 2004. Learning phonotactic distributions. In *Constraints in Phonological Acquisition,* ed. by R. Kager et al. (eds.) Cambridge University Press, pp. 245-291. Yip, M. 2002. *Tone*. Cambridge University Press.

Effects of linguistic structure on perceptual attention given to different speech units

Shinae Kang, Keith Johnson

University of California, Berkeley

sakang2@berkeley.edu, keithjohnson@berkeley.edu

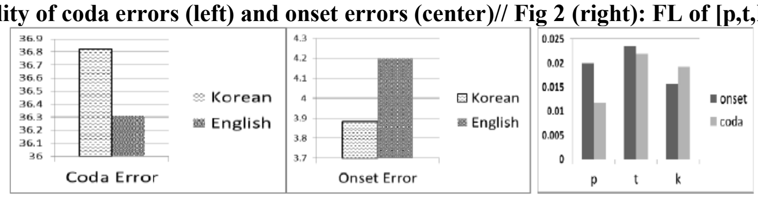
Listeners can shift their attention to speech units of different sizes during speech perception (Goldinger and Azuma 2003).This study extends this claim and investigates if the structure of a language influences the extent to which listeners can give attention to different phonetic units. Lee and Goldrick (2008) found that Korean and English listeners tend to preserve CV- and –VC unit respectively, while shadowing a series of six monosyllables, because of the structural bias of the languages. The primary goal of our study is to see if this structural bias can directly affect perception. The secondary goal is to see if the amount of information that a speech unit conveys also has an effect. Since Korean has a smaller syllable inventory than English, Korean phonemes globally play a smaller functional role than English phonemes (Oh et al. 2013).However, because of the inherent structural bias, the amount of information may differ by syllabic position. We specifically investigate if the amount of information is correlated with the occurrences of assimilation errors by Korean listeners. In short, we focus on the effect of perceptual attention on the perceptibility of intervocalic consonant clusters (VC_1C_2V) and whether it varies cross-linguistically by these structural factors.

Experiment We recorded eight talkers (4-Korean; 4-American English) saying VC- and CV-syllables (vowels: [i, u, a]/ Consonants: [p,t,k]) separately. For each talker, the syllables were spliced back-to-back to create non-overlapping VC_1C_2V-stimuli, where vowels always matched with all combinations of stop clusters. Twenty-nine listeners in the two language groups (15: English/ 14: Korean) participated in a 9-Alternative-Forced-Choice perception task. They identified the intervocalic C_1C_2 as one of the 9 alternatives (“pt”, “pk”, “pp”, etc.).

Analysis We first calculated non-parametric sensitivity for different stop sounds for each listener (I) as a measure of perceptual salience (Hume et al. 1999). Then, the coda (C_1) and the onset (C_2) errors were counted separately for both language groups. For each error type, the number of anticipatory assimilation errors ($C_2 \rightarrow C_1$) and carry-over errors ($C_1 \rightarrow C_2$) were counted respectively in order to distinguish the errors caused by adjacent sounds from others. In addition, we measured the amount of information of [p, t, k] from the Korean spoken corpus by quantifying Functional Load (FL) of each sound based on Shannon-entropy (Oh et al. 2013). For example, FL of [p] is defined as the sum of the amount to which the contrasts are neutralized if [p] and the sound [x] were merged. It is thus associated with both type and token frequency of syllables containing [p].

Result and Discussion While the number of coda errors is greater than onset errors for both language groups, Korean listeners made significantly more errors in coda ($t=-17.4774$, $p<0.01$) and English listeners made significantly more errors in onset ($t=32.59$, $p<0.01$) (Fig. 1). The result is consistent with the finding that Korean and English speakers have different sub-syllabic representation, in that the listeners have more difficulty in the perception of a unit excluded from their pre-disposed chunk (Lee and Goldrick 2008).

Fig 1: The probability of coda errors (left) and onset errors (center)// Fig 2 (right): FL of [p,t,k] in onset and coda



Also, stop sounds have different FL in different positions (Fig. 2). The relative ranking of FL of the coda stops (t>k>p) is strongly correlated with sensitivity for coda stops for each listener. In addition, we also found that FL calculated from different positions and their relative strength is more strongly associated to the actual distribution of [p, t, k] in the tokens that showed assimilation errors than FL calculated from all positions. In summary, the result indicates that the linguistic structure of a language can potentially affect the level of perceptual attention that its users give to a linguistic unit.

References

Goldinger, S. D., & Azuma, T. (2003). Puzzle-solving science: the quixotic quest for units in speech perception. *Journal of Phonetics, 31*(3-4), 305–320.

Hume, E., K. Johnson, M. Seo, G. Tserdanelis & S. Winters. (1999) A Cross-linguistic study of stop place perception. *Proceedings of the 14th International Congress of Phonetic Sciences*. 2069-2072.

Lee, Y., & Goldrick, M. (2008). The emergence of sub-syllabic representations. *Journal of Memory and Language, 59*(2), 155-168.

Oh, Y., Pellegrino, F., Coupé, C. & Marsico, E., (2013), "Cross-language Comparison of Functional Load for Vowels, Consonants, and Tones", *proc. of Interspeech 2013*, Lyon, France, 25-29 August.

The Sejong Spoken Corpus (2009). The National Institute of the Korean Language.

The perception of the tensity contrast in two German varieties

Conceição Cunha & Ulrich Reubold
Institute of Phonetics and Speech Processing (IPS), Munich
{cunha|reubold}@phonetik.uni-muenchen.de

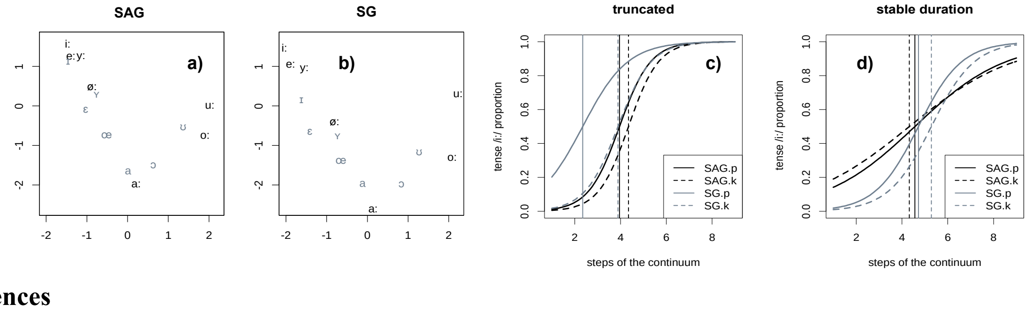
The main aim of this study was to investigate, how the different implementation of the tensity contrast in the production in two varieties of German is also reflected in perception.

German distinguishes tense and lax vowels giving rise to minimal pairs differing only in tensity (e.g. *bieten* 'to offer' vs. *bitten* 'to request'). In Standard German (SG) lax vowels are shorter and less peripheral than their tense counterparts (Fischer-Jørgensen, 1990, Hoole 1999), probably due to the different implementation of tongue acceleration (Hoole and Mooshammer, 2002). However, previous research attested a partial neutralization of the quality contrast in Standard Austrian German (SAG, Moosmüller 2007. Preliminary studies have shown a greater vertical approximation of high and low lax vowels /ɪ, ʏ, ʊ, a/ in the direction of the tense counterparts /i, y, u, a/ for SAG compared to SG, but also a comparable vowel quantity contrast.

For SG, we expect vowel quality to be more important on the discrimination of tense and lax vowels, since this is the distinctive contrast in German and the vowel quantity may only be a consequence of the reduced trajectory of the lax vowels (Hoole 1999). Since SAG neutralizes the quality but not the quantity vowel contrast in production, we predict that vowel quantity may be a more important cue in perception for SAG speakers than for SG speakers. An additional aim was to test, whether consonantal context (symmetrical /p_p/ vs. /k_k/) may also influence the category boundary, as the tensity contrast in SG high front vowels has been shown to be smallest in velar context and comparable to the same contrast in SAG (Harrington et al., 2012). This results in two hypotheses: **H1**: In the *truncated* continuum, SG listeners should perceive more tense [i] and present a left shifted category boundary than SAG listeners, but only in bilabial context. **H2**: SG listeners will show categorical perception in the *stable duration* continuum, but such a clear distinction should be difficult for SAG listeners, which should result in flatter slopes in their response curves.

In order to test these hypotheses, we prepared two 9-step continua from tense /i/ to lax /ɪ/: 1.: we truncated a tense vowel (F1: 220 Hz, F2: 2100 Hz) from its offset per step by 10% of the original length of 108 ms (resulting in a continuum from 20% ('lax', 21.6 ms) to 100% ('tense'), while maintaining the formant values of the tense vowel. 2.: we maintained the duration of the middle step of the first continuum (step 5 with 60 % (64.8 ms) of the original tense vowel length) and manipulated only the first two formants from the 'tense' values reported above to the speaker's approximate mean values for the lax [ɪ]s (F1: 320 Hz, F2: 1600 Hz) in nine bark sized steps.; we embedded these two vowel continua into /p_p/ and /k_k/ contexts and presented the resulting logatomes to 21 SAG and 24 SG subjects. For the statistics we ran ANOVAS on the category boundary, slope and intercept with consonant and variety as factors separately for both continua and post hoc t-tests on each combination of consonant and dialect.
SG listeners' category boundary to the *truncated* continuum was left shifted in bilabial context in comparison to all other boundaries ($p<0.001$ in all three cases), whereas all other pair-wise comparisons revealed insignificant results. The category boundaries did not differ between *stable duration* continua, but considering the slopes we found a main effect of variety ($F[1,41]=10.9$, $p< 0.01$), in which slopes were flatter for SAG ($p<0.001$ for both contexts) than for SG.
This study has shown the use of different perceptual cues resulting from different realizations of the same phonemic contrast. At least for the tense high front vowels, the tensity contrast in SAG is suggested to be implemented rather as a quantity contrast in both perception and production.

Figure 5: PCA-transformed tongue data extracted at the temporal midpoint of several tense-lax vowel pairs in **a)** Standard Austrian German (SAG) and **b)** Standard German (SG); response curves and 50% perceptual boundaries to the **c)** *truncated* and **d)** *stable duration* /i/-/ɪ/-continua, embedded into velar (dashed) and bilabial (solid) contexts, for SAG (black) and SG (grey) listeners.



References

Fischer-Jørgensen, E. (1990): Intrinsic F0 in tense and lax vowels with special reference to German. *Phonetica* 47, 99-140. / Harrington, J., Hoole, P., Reubold, U. 2012. A physiological analysis of high front, tense-lax vowel pairs in Standard Austrian and Standard German. *Italian Journal of Linguistics* 24, 158-183. / Hoole, P. 1999. *On the lingual organization of the German vowel system. J. Acoust. Soc. Am.* 106, 1020-1032 / Hoole, P., Mooshammer, C. 2002. Articulatory analysis of the German vowel system. In P. Auer, P. Gilles & H. Spiekermann (eds.), *Silbenschnitt und Tonakzente*. Tübingen: Niemeyer, 129-152. / Moosmüller, S. 2007. *Vowels in Standard Austrian German*. Vienna: University of Vienna.

Mismatch Negativity Reveals Abstract, Monovalent Phonological Primitives

Mirjam J.I. de Jonge
University of Amsterdam
m.j.i.dejonge@uva.nl

Background. A long-standing issue in phonological theory is what the primitives of phonological representations are. Two competing theories are Element Theory (Backley, 2011), positing *elements* as representational primitives, and the Featurally Underspecified Lexicon model (FUL; Lahiri & Reetz, 2002), positing *features*. The Mismatch Negativity (MMN; Näätänen, 2001) is a pre-attentive brain response to acoustic change that is also sensitive to phonological factors. We use the MMN to evaluate predictions of perceptual responses arising from Feature-based versus Element-based representations. Of interest are the level of *granularity* that is seen in the neural responses – i.e. to what extent MMN amplitudes gradiently reflect acoustic distances rather than abstract categorical features – and the *directionality* of responses within a contrast – i.e. whether the response is stronger in one direction than the other. If all vowel properties are represented with equal weight, then MMN magnitude must be similar for each direction of a contrast. However, if all properties are not represented equally, as is by definition the case with monovalent features, one direction of contrast is expected to be stronger than the other. According to FUL, changes from richer to less specified stimuli give the strongest mismatch and thus a stronger response. Since Element representations specify the oppositive properties from FUL (see table 1), the predicted asymmetries are also reversed.

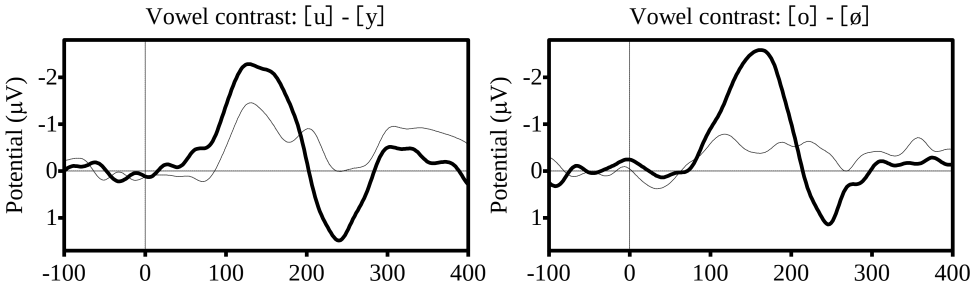
Method. Synthesised tokens of the French vowels /y/, /u/, /o/ and /ø/ (see Table 1; duration 150 ms) were presented in a four-block triple-deviant oddball paradigm, each vowel serving as standard (p=.85) in one block and as deviant (p=.05) in the three other blocks. Participants were 24 righthanded monolingual native speakers of French, their EEG was recorded during passive listening while they watched a silent subtitled movie. MMN was computed by subtracting the grand average ERP of a vowel as a standard from the grand average ERP of the same vowel as a deviant.

Findings. Preliminary analyses of the first 20 participants (see fig. 1) show granular and asymmetric response patterns, indicating that listeners parse speech sounds into abstract and monovalent units of information. The granularity is apparent from the similarity between the MMN of [ø]-deviants in [o]-standard sequences and the MMN of [y]-deviants in [u]-standard sequences (solid lines in fig.1). While the F2 change in the vowels pairs is very different (see table 1), the MMN is not, reflecting the abstract phonological property that is involved in both contrasts rather than the acoustic distance. The asymmetry shows up as larger MMN for changes from a back standard to a front deviant and smaller MMN for the opposite change. This is in line with the FUL model, in which back vowels are [DORSAL] and specified as such, but front vowels are [CORONAL] and therefore underspecified for place of articulation. The direction of the detected asymmetry runs counter to the prediction based on representations with Elements.

Table 1: Vowel formants and their representations according to Element Theory and FUL.

Formants		Representations	Representations		Formants
F1: 308 Hz F2: 1750 Hz	**y**	[HIGH] [LAB] [COR] | I U |	[HIGH] [LAB] [DOR] | U |	**u**	F1: 308 Hz F2: 764 Hz
F1: 376 Hz F2: 1417 Hz	**ø**	[LAB] [COR] | I U A |	[LAB] [DOR] | U A |	**o**	F1: 376 Hz F2: 793 Hz

Figure 1: Difference waves at electrode position Fz. Solid line: [y,ø] deviant. Dotted line: [u,o] deviant.



References

Backley, P. 2011. *An Introduction to Element Theory*. Edinburgh: Edinburgh University Press. / Lahiri, A. & Reetz, H. 2002. Underspecified recognition. In: Gussenhoven, C., Warner, N. & Rietveld, T. (eds.), *Labphon 7*. Berlin: Mouton, 637-676. / Näätänen R. 2001. The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent MMNm. *Psychophysiology* 38, 1–21.

Phonotactic learning and its interaction with speech segmentation

Alexis Black and Masaki Noguchi
University of British Columbia
akblack2g@gmail.com; mskngch@gmail.com

Productive acquisition of phonotactic regularities is clearly of crucial importance for the language user; however, it is also possible that these regularities serve a facilitative and functional role in the language learning process. One hypothesis along these lines is that phonotactic rules can accelerate/bootstrap the development of the lexicon (Brent & Cartwright, 1996; Thiessen & Saffran, 2003). Phonotactic rules rely on abstract phonological representations, such as phonological features and natural classes. While it is now well established that adult listeners can segment continuous speech based on the transitional probabilities (TPs) between adjacent syllables and non-adjacent segments (c.f. Romberg and Saffran, 2010), the question of whether adult listeners have access to abstract phonological representations as they track TPs remains unresolved. In this study, we probe this question by exposing participants to a continuous stream of unfamiliar speech that is structured both with syllable-based TPs as well as a [+/- back] vowel harmony system. We hypothesized that if learners are able to track both phonetic and abstract phonological features, they will experience an enhanced ability to segment the speech stream. There are two alternatives to this scenario: 1) participants may track both TPs and vowel features, but experience a decline in segmentation performance as compared to a control group, due to memory overload; or 2) learners may fail to track the abstract phonological feature, and therefore be unaffected by the harmony system.

17 participants were exposed to a 2-min continuous speech-stream that consisted of 4 trisyllabic TP-defined words (syllable TPs within a word=1.0; across word boundaries=.33). All vowels in the word-string share the feature [-back] or [+back]. Participants were tested on their knowledge of the harmony system and the word strings via two blocks of 2-AFC-questions. In one block, words, part-words and fake-words were pitted against each other to probe participant's knowledge of the words. Part-words are trisyllabic strings that occurred in the speech stream, but that do not line up with word-internal 1.00 TPs. Fake-words, on the other hand, are strings in which at least one of the two syllable transitions never occurred in the speech stream (i.e. a TP of 0.0); however, all syllables in the string occupy the positions that they belong to in the TP-defined words of the language. For example, given words of type ABC and DEF, part-words are of the type CDE, whereas fake-words are of the type AEF. Part-words and fake-words are further divided into harmonic and non-harmonic types, providing an indirect test for knowledge of the harmony system. A second block of 2-AFC items tests participants more directly on their knowledge of the harmony system by pitting non-words – trisyllabic sequences of completely novel syllables – against each other. The order of block presentation (i.e. harmony test vs word test) was counterbalanced across participants. 20 participants were run in a control condition, in which the language lacked vowel harmony.

The results show that participants successfully segment the language based on TPs (M=67% accuracy, $t(16)=4.78$, $p<.001$); this performance is furthermore equivalent to the control condition (M=68%, $t(35)=.-.23$, $p=.82$). There is little evidence that participants have learned the harmony system. Performance on the direct harmony tests is at chance level (M=52%, $t(16)=0.02$, $p=.985$). In the indirect harmony tests, performance is roughly equivalent across all contrasts, independent of the foil's harmonic structure. However, there is one condition in which harmonic/non-harmonic choices elicited differential responses. When participants were asked to choose between a non-harmonic fake-word of type 1 (see table) and non-harmonic part-word, they were significantly more likely to choose the fake word. When the fake- and part-word pair were harmonic, however, participants were at chance performance. This finding suggests that participants may have acquired some aspect of the harmonic system. Future work will explore this possibility by manipulating the input language. For example, a larger number of words would increase the complexity of the segmentation task, but perhaps increase participants' likelihood of forming generalizations.

Table 1. 2-AFC item types.

Words	**Part-Words**		**Fake-Words**			**Non-Words**	
	Harmonic	*Non-Harmonic*	Type	*Harmonic*	*Non-Harmonic*	*Harmonic*	*Non-Harmonic*
budoga gilebae	dogata	dogapae	1	**ta**doga	**gi**doga	laegebi	tiralae
tapuro paedite	ditegi	diteta	2	bu**pu**ga	gi**do**bae	gotura	gopebi
			3	budo**ro**	gile**ga**		

References

Brent, M. R., & Cartwright, T. A. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition*, *61*(1), 93-125. / Romberg, A. R. & Saffran, J. R. (2010). Statistical Learning and Language Acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*. *1*(6), 906-914. / Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: use of stress and statistical cues to word boundaries by 7-to 9-month-old infants. *Developmental psychology*, *39*(4), 706.

Effects of a Formant Transition of Preceding Vowel off-glide on Perception of Japanese Geminate Consonant *Sokuon*

Emi Yanagisawa, Takayuki Arai

Sophia University

emiyana@hotmail.com, arai@sophia.ac.jp

The primary acoustic correlate of a single-geminate distinction in Japanese is a difference in duration－closure duration for stops and frication duration for fricatives (Kawahara, in press). As durational cues, not only absolute duration of a geminate consonant but also relational correlates such as ratio of stop closure to the preceding vowel, following vowel, preceding mora or word are known as cues for single-geminate distinction. Non-durational cues of a single-geminate contrast in Japanese such as F0, intensity and voice quality are also investigated (Idemaru and Guion, 2008). Nakano (1974) indicates that an abrupt damping in intensity of preceding vowel off-glide is associated with a perception of Japanese obstruent geminates “Sokuon”. We assume that the damping is caused by a contact or approximation of the articulator in order to articulate obstruent or fricative consonants and a formant transition ought to be associated with the damping. Yanagisawa and Arai (2013) conducts perception experiment using natural speech and reports that an utterance does not tend to be perceived as Sokuon when an off-glide of preceding vowel does not include formant transition and steep intensity damping. According to Yanagisawa and Arai (2013), the formant transition seems to play one of the roles as a perceptual cue of Sokuon. However, it is not confirmed as yet by any researches. Therefore, the present study tested the effects of formant transition of preceding vowel on Sokuon perception systematically.

The present study conducted a perception experiment using synthetic sounds, non-sense disyllables /V_1CV_2/ (/ata/-/atta/), as stimuli. Two transitional patterns of the first and second formants, F1 and F2, of /V_1/ were used to synthesize the stimuli, i.e., with or without formant transition (FT1 or FT0, respectively). The closure duration of /C/ was gradually lengthened in 15 levels and the damping of intensity was not included in /V_1/ in order to eliminate the effect from the intensity damping (see Fig.1). The stimuli were repeated three times and randomized when they were presented to participants. 25 native Japanese listened to the stimuli and evaluated them either singleton or geminate consonant.

Figure 2 shows two sigmoidal functions of the proportion of geminate stop responses as function of the /C/ closure duration, each of which corresponds to one of the two formant transitional patterns. There are two main findings from the present study. First, geminate responses increase as closure duration increase. It supports previous studies that claim the closure duration is the primary cue of Sokuon perception. Second, FT1 utterances are perceived as Sokuon with shorter /C/ closure duration (see Fig.2). The cross-over point is 205 ms on FT1 and 248 ms on FT0. Furthermore, from the observation of geminate responses on Fig. 2, when the response is 50% on FT1, the response on FT0 is 21% by only having removed the formant transition from the FT1 utterance. And FT0 utterances tend not to be perceived as Sokuon even though the /C/ closure duration is 380 ms (the longest duration in the present study). This tendency is also reported in Yanagisawa and Arai (2013) using a natural speech. It is presumably because the formant transitions reflect the acoustical characteristics due to the tongue movement at the end of the preceding vowel and are necessary for Sokuon whereas it is not the case for the non-Sokuon context. Sokuon is assumed to be perceived distinctly when the formant transition is accompanied along with the durational cues. From these discussions above, it can be concluded that a formant transition affects Sokuon perception as one of the non-durational cues.

Figure 1: Parameter changes in time for the formant synthesis

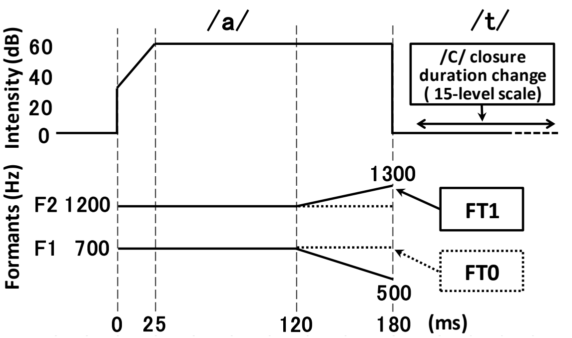
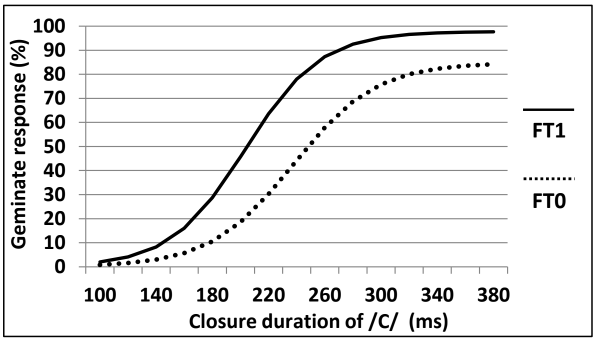


Figure 2: % response of geminates vs. /C/ closure duration with and without formant transitions



References

Idemaru, K., and Guion, S. 2008. Acoustic covariants of length contrast in Japanese stops. *Journal of the International Phonetics Association*. 38, 2, 167-186. /Kawahara, S. in press. The phonetics of obstruent geminates, sokuon. *The Handbook of Japanese Language and Linguistics*. Phonetics and Phonology: Mouton. /Nakano, K. 1974. The phonetic features of the choked sound in Japanese. *Onseigaku sekai ronbunshu*. 719-724. /Yanagisawa, E., and Arai, T. 2013. Acoustic characteristics of vowel off-glide followed by Japanese geminate. *2013 Spring Meeting Acoustical Society of Japan*. 413-414.

An empirical assessment of the carryover bias in tonal coarticulation

Feng-fan Hsieh
National Tsing Hua University
ffhsieh@mx.nthu.edu.tw

Introduction: This study is an empirical assessment of two competing accounts of the carryover bias in tonal coarticulation (as well as the rightward bias of tone spreading (Hyman & Schuh 1974, Hyman 2007, *a.o.*)). Xu & Liu (2006) and subsequent papers hypothesize that syllables are coupled with pitch targets/tones in full synchrony (i.e., the in-phase mode) for the sake of stability in speech production. Since speech is linear in nature, it follows that the carryover bias is motivated since the pitch target/tone at the end of the preceding syllable must be faithfully rendered *before* the pitch targets of the following syllable are realized. A similar proposal is found in Akinlabi & Liberman's (2001) "Tonal Complex". On the other hand, Flemming (2011) proposes a perceptual account, namely that since i) the fact that sonorous rimes contain well-structured harmonics greatly facilitates tonal discrimination, ii) therefore, vowels provide "best platform for the realization of tone," followed by sonorants and subsequently obstruents (Gordon 2004, *a.o.*) and iii) Zhang (2004) concludes from his cross-linguistic survey that the duration of the sonorous rime is a vital factor in licensing contour tones, so it is predicted that in sequences like $CV_1.CV_2$, the faithful realization of tone should be prioritized in V_1, but not in the interlude C. To this end, the two theories make the exactly same prediction about the bias in question. However, Flemming's (2011) analysis further predicts "syllables with obstruent codas should pattern differently from sonorant codas." Early transitions with obstruent codas are expected since F_0 cannot be realized on obstruents. By contrast, under Xu & Liu's (2006) view, obstruent and sonorant codas should pattern alike with respect to coarticulatory effects simply because they are both syllabified as the coda (i.e., in the same prosodic position).

Goal: We believe that Taiwanese (Southern Min) provides a good testing ground for the two competing theories. This is because syllabic nasals (ŋ̍ and m̩) are phonemic in this language. Therefore, in sequences like (C)N̩.V (i.e. a syllabic nasal precedes an onsetless syllable), tonal coarticulatory patterns should be more or less the same as those found in CV.CV sequences, according to Xu & Liu's (2006) account, whereas Flemming's (2011) approach predicts that more anticipatory coarticulatory effects should be found in (C)N̩.V. Recall that vowel priority in tonal faithfulness is invoked to derive the carryover bias. In other words, if there is *no* vowel in the preceding syllable, then early transitions to a following (unlike) tone are expected (i.e. more anticipatory coarticulatory effects).

Method: Eight (8) native speakers of Taiwanese were recruited to participate in this experiment (3 male and 5 female; aged 22 to 28; all are from Southern Taiwan). All target phrases are composed of either $(C_{[son]})$N̩.V or $C_{[son]}$V.NV and Vowels are controlled for height in NV.NV sequences (to minimize the effect of vowel intrinsic F_0). All of the test phrases are meaningful, e.g., (C)N̩.V [nŋ̍^{H}.i^F] 'soft chair' vs. CV.NV [lu^H.mi^H] 'woman noodle,' or [le^L.ŋɔ^F] 'Example 5'. There are four (4) sandhi tones in sandhi-initial position: {H, M, L, F} and five (5) tones in sandhi-final position: {H, M, L, F, low R}. Sandhi-initial and final tones are completely neutralized, i.e., sandhi-initial H, for example, is not distinguishable from H in final position, etc. (see Myers & Tsay 2008). All stimuli were randomized and embedded into a carrier phrase: "I transcribe ___ transcribe three times", where the target words were flanked by mid tone words [tsʰau^{33}] 'transcribe'. Each carrier phrase was repeated six (6) times, yielding a total number of 3,840 tokens (= 4 sandhi-initial tones × 5 sandhi-final tones × 2 combinations × 2 types of vowel height (H vs. non-H) × 6 repetitions × 8 speakers). We used SS ANOVA (Gu 2002, see also Davidson 2006) to quantify the difference of F_0 contours.

Results: The experimental results are summarized below. When a tone is marked out for both combinations, that means the anticipatory effects are equivalent. Otherwise, M1's L in CN̩.V, for example, means that more anticipatory effects are found in sandhi-initial L in CN̩.V than in CV.NV. Note again that these are all based on SS ANOVA results.

	M1	M2	M3	F1	F2	F3	F4	F5
(C)***N̩***.V	<u>L</u>	H/<u>L</u>	H	H/M/F	M/<u>L</u>	<u>L</u>/F	H/M/<u>L</u>/F	<u>L</u>/F
C***V***.NV	H/M/<u>F</u>	H/<u>F</u>	H/L/<u>F</u>	H/M/L/<u>F</u>	<u>F</u>	<u>F</u>	<u>F</u>	N/A

Discussion: Our results show that neither is confirmed. Male speakers tend to allow more coarticulatory effects in CV.CV, while the magnitude of tonal coarticulation is more pronounced on female speakers' syllabic nasals. Despite this gender difference and inter-speaker variation, some "tone-specific" tendencies can still be found, namely that L on syllabic nasals is more subject to contextual influence and F on vowels undergoes more anticipatory coarticulation. Notably, the above observations are not confounded with vowel height, duration of vowel/syllabic nasal, tonal space, age, (sub-)dialects and so on. Some possible explanations will be discussed during the presentation.

Selected References Davidson, L. 2006. Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *J. Acoust. Soc. Am.* 120, 407-415. Flemming, E. 2011. La grammaire de la coarticulation. In: Embarki, M., Christelle, D. (eds.), *La coarticulation: Des indices à la représentation*. Paris: L'Hartmattan, 189-211. Xu, Y., Liu, F. 2006. Tonal alignment, syllable structure and coarticulation: Toward an integrated model. *Italian Journal of Linguistics* 18, 125-159.

Swedish focal accents and the syntax–prosody interface

Sara Myrberg
Stockholm University, Department of Swedish Language and Multilingualism
sara.myrberg@su.se

Introduction: This paper discusses the status of the *focal accent* (FA) in Stockholm Swedish as a *marker of phonological phrasing*. Swedish and Norwegian are well-known for their lexical tone accent distinction (accent 1 vs. 2), but also for a categorical distinction between two types of tonal prominences: word accents (have been considered the realization of lexical tones), and focal accents (post-lexical tone) (Bruce 1977, Fretheim & Nilsen 1989). The term FA refers to its correlation with information structural focus, but Bruce originally termed it 'sentence accent', alluding to it being (nearly) obligatory in a sentence. However, the details of FA distribution have not been fully understood. In well-described West Germanic languages (e.g. Pierrehumbert 1980), *pre-nuclear* and *nuclear* pitch accents have been distinguished. Though there is no full consensus, researchers seem to agree that nuclear accents correlate strongly with focus, and appear in (nearly) all sentences. It may be appealing, then, to think of the FA as the nuclear accent (or head of the Intonation Phrase) in Swedish. This paper, however, provides new data suggesting that the FA is better described as the head of the Phonological Phrase (Myrberg 2010). FAs primarily correlate with syntactic XPs, rather than with sentences or clauses (as may be expected of a nuclear accent or head of Intonation Phrase, cf. Selkirk 2009, Itô & Mester 2011). It follows that FAs can be nuclear or pre-nuclear, and that FA distribution should conform to the OT-constraint STRESSXP (Truckenbrodt 1995). This discussion is intended to shed light on the elusive relation between the West Gmc nuclear ~ prenuclear distinction vs. the Swedish and Norwegian (North Gmc) Word accent ~ FA distinction.

Data: A corpus of 1200 read sentences is presented, illustrating the distribution of focal accents in sentence initial noun phrase (NP) subjects, immediately preceding the finite verb (Swedish is a V2 language). Subject NPs appear with 4 different lengths: 2 words (*brown hare*), 3 words (*brown hare with kids*), 4 words (*brown hare with many kids*), 5 words (*brown hare with many cute kids*) (note: the preposition 'with' does not form a prosodic word). All subjects consist of initially stressed disyllabic words. There are 5 items with lexical accent 1, and 5 with lexical accent 2. Each item is read once with narrow focus on the subject (the verb phrase is given) and once with narrow focus on the verb phrase (the subject is given). Sentences are repeated 3 times, randomized and interspersed with filler sentences. 5 native Stockholm Swedish speakers were recorded (total: 4*(5+5)*2*3*5=1200). The data was annotated semi-automatically using Praat (Boersma & Weenink). In each target word, tonal points were automatically inserted (followed by manual correction). In addition, the author made (subjective) judgments with respect to whether each word had a FA. The judgments are independent of the tonal annotation, and can therefore be evaluated via this annotation. In this abstract, data with lexical accent 2 is presented (600 sentences), pending analysis of lexical accent 1 data.

Results: As expected, the focused constituent (subj or verb phrase) always had a FA on its rightmost word. However, there are additional FAs. **When the subj was given**, 29,3% (22/75) of 2-word subj nonetheless had a FA on the rightmost word (*brown* ***hare***), and 55% (41/75), 52% (39/75), and 50% (38/75) of respectively 3-, 4-, and 5-word subjects had a FA on the final word. Long subjects (4 and 5 words), frequently had multiple FAs. Among 4-word subjects with a final FA, 56% (22/39) had a second FA (*brown* ***hare*** *with many* ***kids***). Among 5-word subjects with a final FA, 68% (26/38) had a second FA (*brown* ***hare*** *with many cute* ***kids***). The second FA always appeared on the head noun (*hare* in this example). 2-word subjects never had two focal accent and two subjects only rarely did (7 out of 41 times). **When the subject was focused**, longer subjects also sometimes carry FA in addition to the final FA marking focus: 27% (20/73) of 5-word subjects and 19% (14/75) of 4-word subjects had such additional FAs.

Discussion: I argue that the frequent occurrence of prenuclear FA suggests that FA is not primarily a nuclear accent (one per clause), but that it is better analyzed a marker of the Phonological Phrase (one per XP). I also discuss, however, the fact that the STRESSXP (Truckenbrodt 1995) constraint would predict even more FAs, indicating that Swedish is restrictive in terms of PP-phrasing. In addition, the height of FA in prenuclear vs. nuclear positions will be measured, to give an indication with respect to whether nuclear FAs are phonetically distinguished from pre-nuclear ones.

References: Bruce, G. 1977. *Swedish word accents in sentence perspective*. Lund: Liber Läromedel. / Fretheim, T. & Nilsen R. A. 1989. Terminal rise and rise-fall tunes in East Norwegian intonation. *Nordic Jnl of Ling* 12:155–181. / Itô, J., & Mester, A. 2011. Recursive prosodic phrasing in Japanese. In: T. Borowsky, S. Kawahara & M. Sugahara (eds.), *Prosody matters: Essays in honor of E. Selkirk*. Equinox, London. / Myrberg, S. 2010. *The Intonational Phonology of Stockholm Swedish*. ACTA Universitatis Stockholmiensis 53. / Pierrehumbert, J. 1980. *The phonology and phonetics of English intonation*. PhD thesis, MIT. / Selkirk, E. 2011. The syntax–phonology interface. In: J. Goldsmith, J. Riggle, & A. Yu (eds.) *The handbook of phonological theory, 2nd edition*. Blackwell, Oxford. / Truckenbrodt, H. 1995. *Phonological Phrases: Their Relation to Syntax, Focus, and Prominence*. PhD Thesis, MIT.

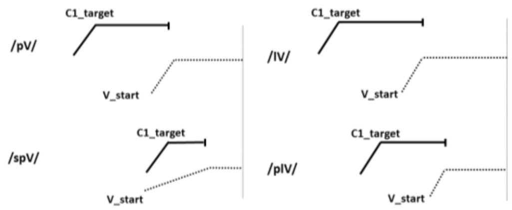
Vowel timing in complex onsets of American English

Stavroula Sotiropoulou
Department of Linguistics, University of Potsdam
stavroula.sot@gmail.com

Temporal stability patterns as a function of syllable position have been hypothesized to emerge from certain coupling relations among segments in the syllable. Previous studies have shown that English syllable initial clusters are timed with the vowel in the pattern known as the C-center effect (Browman & Goldstein, 1988; Marin & Pouplier, 2010). The C-center effect results from competitive coupling among segments in the syllable. Being the onset of the syllable, both consonants of the cluster should start synchronously with the vowel. At the same time for matters of perceptual recoverability, the two consonants should be sequential to each other. As a result of this, it is the consonant center of the cluster (C-center) which seems to maintain a stable relation to the vowel. A prerequisite for the C-center organization is the ability of the prevocalic consonant to overlap with the vowel. However, in an analysis of X-Ray microbeam data, Goldstein et al. (2009) report that the C-center effect is dependent on segmental properties of the consonants that partake in the syllable. The dependence shows up most clearly between /pl/ and /sp/ onsets concerning the ability of the prevocalic consonant (C) to overlap with the vowel (V). By examining the distance in time between the targets of C and V (henceforth, "CV lag" or simply "lag"), /p/ seems to overlap substantially more with the vowel when /s/ is added to the syllable (i.e. the lag between /p/ and /V/ decreases in /spV/ compared to /pV/), while /l/ resists overlap when /p/ is added (i.e. minimal decrease in lag in /plV/ compared to /lV/). None of the past studies has examined the vowel initiation in clusters and how the vowel gesture may change when a consonant is added to the syllable. Shortening of the V driven by the addition of a consonant in the syllable is a necessary but not a sufficient condition for C-center organization. For example, vowel shortening in itself does not necessarily indicate C-center patterning (Katz, 2012). The V can shorten without the V starting at the C-center of the cluster.

Using Electromagnetic Articulography, our aim is to assess whether the results of Goldstein et al. (2009) can be replicated and to elaborate on any effects using more vocalic contexts and converging measurements. We also report, for the first time, results on word-medial clusters, which we expect to pattern as in word-initial clusters since the syllabification is presumed to be the same across the two positions (e.g. spend and suspend). In addition, we report, also for the first time, on measures of vowel articulatory initiation across CV and CCV, which more directly assesses if and to what extent the prevocalic consonant overlaps with the vowel. Examining different landmarks of the vowel enables us to form a thorough picture of how the vowel unfolds and how its gesture changes from singleton to cluster. We report here results from three subjects (two more will be added by the time of the meeting). Our results indicate a larger decrease in CV lag from /pV/ to /spV/ than from /lV/ to /plV/ both in initial (pMCMC= .000) and medial position (pMCMC= .006). This both replicates earlier findings and extends these to different vowel heights and the word-medial context. By examining different landmarks, we find that in /spV/ the vowel gesture shortens non-uniformly: the plateau of the vowel gesture seems to endure shortening while the to-target part of the gesture seems to bear elongation in a way that it overlaps largely with the prevocalic /p/ (see Figure 1, left). The decrease in lag between /lV/ and /plV/, instead, occurs by a uniform shortening of the whole vowel gesture and without an earlier start of the vowel as we see in /spV/ (see Figure 1, right). Therefore, the internal temporal structure of the vowel gesture is different across /sp/ and /pl/.

Figure 6: Vowel initiation patterns across /pV/~/spV/ and /lV/~/plV/.



In conclusion, we provide further evidence that complex onset clusters in English have cluster-specific timing patterns. We furthermore quantify these timing patterns also by vowel initiation measures for the first time. The timing patterns are strongly linked to the internal temporal structure of the vowel gesture. Different types of vowel compression in complex onsets are effected via different adjustments to the internal temporal structure of the vowel. What was previously referred to as vowel shortening or compression is not articulatorily a unitary concept.

References

Browman, C., Goldstein, L. 1988. Some notes on syllable structure in articulatory phonology. *Phonetica* 45, 140-155. / Goldstein, L., Nam, H., Saltzman, E., Chitoran, I. 2009. Coupled oscillator planning model of speech timing and syllable structure. In: Fant, G., Fujisaki, H., Shen, J. (eds.), *Frontiers in Phonetics and Speech Science*. Beijing: The Commercial Press, 239-250. / Katz, J. 2012. Compression effects in English. *Journal of Phonetics* 40 (3), 390-402. / Marin, S., Pouplier, M. 2010. Temporal organization of complex onsets and codas in American English: Testing the predictions of a gestural coupling model. *Motor Con*MIT.

The processing of schwa reduction in isolation and in connected speech: An electrophysiological study

Kimberley Mulder and Linda Drijvers

Radboud University, Centre for Language Studies, Nijmegen, The Netherlands

Kimberley.Mulder@mpi.nl

In spontaneous speech, speakers often do not produce words in their fully articulated form. Rather, words occur often in a reduced form, with altered or fewer speech sounds and even with fewer syllables. Examples of reduced pronunciation variants are the English words *s'pose* and *yeshay* for *suppose* and *yesterday*, respectively. Native listeners are generally not aware of the presence of these reduced pronunciation variants and understand these variants effortlessly. Nevertheless, recent behavioral studies suggest that unreduced forms benefit from a privileged status: The studies report a processing advantage for the unreduced variant of words over reduced variants, especially when these words occur in isolation (e.g., Ernestus and Baayen, 2007). The question is whether this processing advantage for unreduced pronunciation variants remains when words are presented in a more natural context, namely in natural connected speech.

In the present study, we investigate the processing of reduced and unreduced pronunciation variants using electrophysiological measures. These measures are highly sensitive and allow us to study words in connected speech without requiring listeners to make artificial decisions about words (e.g., lexical decision, semantic categorization). We investigated the brain response to reduced and unreduced words in two different listening contexts: in isolation (Experiment 1), and in a full sentence at sentence-final position (Experiment 2). In the two experiments, native speakers of Dutch listened passively to 200 Dutch stimuli while their brain activity was recorded, and they made comprehension questions after blocks of stimuli. The reduction phenomenon under investigation was schwa reduction in the first syllable of Dutch prefixed infinitives, such as *genieten* ('to enjoy'). Schwa reduction is a subtle type of reduction, and the absence of a schwa in prefixed infinitives is not easily detectable during listening, even when schwa reduced infinitives are presented in isolation. This type of reduction is therefore extremely suitable for comparing the processing of reduced words in isolation and in sentences. Target words were 80 infinitives realized without a schwa in the prefix, and the same 80 infinitives pronounced in their full form. Participants heard both reduced and unreduced target infinitives but only heard one version of each infinitive.

Schwa reduction was expected to affect the N400, an ERP component reflecting difficulty of lexical-semantic activation and integration within a given context. We expected that the processing advantage for unreduced variants in isolation as observed in the behavioral literature would translate into smaller N400s for these unreduced variants compared to reduced variants of the same word. However, in sentences (Experiment 2), a context in which segment reduction occurs frequently, this advantage may be lost. The results show that both in isolation (Experiment 1) and in sentence context (Experiment 2) unreduced variants in isolation elicited a smaller N400 compared to reduced variants. The effect appeared to be equally large in both experiments. In addition, we observed in both experiments, early effects of schwa reduction, emerging around 100 ms in the EEG signal.

Our findings show that, during word processing, listeners are sensitive to segment reduction, even when the reduction process is fairly subtle. The early effects in the EEG signal suggest that they are able to detect the absence of a schwa even before the target word has been fully heard. Our data support the findings from behavioral studies that suggest that unreduced variants of words presented in isolation benefit from a processing advantage over the reduced variants of these words. Interestingly, the unreduced variants were also easier to process when they occurred in sentence context at sentence-final position. Apparently, although reduced variants are highly frequent in connected speech, the unreduced variant of prefixed infinitives acts as the more natural variant for this context. In terms of processing, this would imply that, in general, for full forms the route from acoustic input to the activation of their semantic representation in the lexicon is faster. The question is where on this route to activation, the process is hindered for reduced words. The larger N400 for reduced words suggests that lexical-semantic activation is more difficult when segments are deleted. In addition, the early effects suggest that initial segment processing is also affected by the schwa reduction. In conclusion, our findings show that even subtle differences in the realization of spoken words are picked up by the brain and can affect the processing of these words.

References

Ernestus, M., & Baayen, R. H. (2007). The comprehension of acoustically reduced morphologically complex words: The roles of deletion, duration and frequency of occurrence. In Proceedings of ICPHS, Saarbrücken, Germany (pp. 773–776).

Applying phonological knowledge to phonetic accommodation

Emily Clare
University of Toronto
e.clare@mail.utoronto.ca

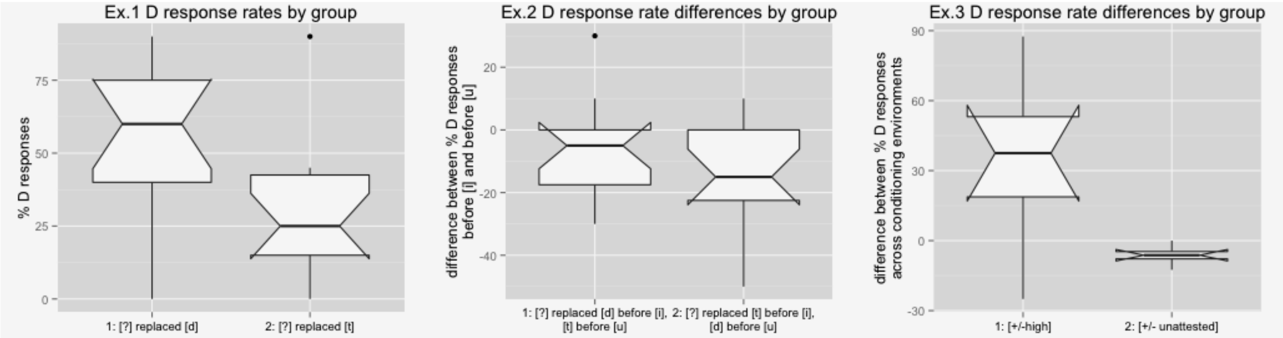
In a groundbreaking study of perceptual learning, Norris et al. (2003) found that listeners shift the boundary between two phonemes along a perceptual continuum following brief exposure to speech exhibiting that pattern. Further studies, such as Kraljic and Samuel (2006) and Nielsen (2011), have tested whether participants would extend their knowledge about the continuum they were exposed to (e.g. [d]-[t]) on the basis of natural classes; listeners indeed shifted the boundary of a related continuum (e.g. [b]-[p]) that was absent from exposure. Although it is tempting to attribute this to speakers' knowledge of phonological features, it is not clear that this is the case. The difference between [d] and [t] acoustically is the same as the difference between [b] and [p], which creates a confound; participants could have been generalising across acoustic traits rather than accessing phonological categories. In this paper, I present results from a new methodology designed to circumvent this issue, which show that natural classes can aid perceptual learning.

Experiment 1 replicated past studies, where participants shifted their category boundary after listening to speech including an ambiguous sound [?] which was consistently disambiguated via lexical knowledge. For example, an ambiguous sound [?] between [d] and [t] in the word *dash* ([?]ash) would be categorised as D because *tash* is not a word. Hearing [?] as [d] caused listeners' VOT boundary to increase (from 30ms to 40ms, for example), as evidenced by their subsequent categorisation of that same [?] in nonsense syllables as D. The rate of D responses by group is shown in Figure 1 (left). A mixed-effects logistic regression model indicated that the main effect of group is significant, showing that participants are adjusting their VOT boundary based on exposure, as has been shown before.

Experiment 2 introduced second-order learning (along the lines of Onishi et al., 2002), whereby the direction of the VOT boundary shift was dependent on the neighbouring vowel. For example, one group of listeners heard [?] in place of [t] before [i] and [?] in place of [d] before [u]. At test, these participants more often identified the ambiguous [?] token as D before [i], but as T before [u], following their exposure. This shows that listeners are able to make perceptual adjustments which are conditioned by a particular adjacent vowel. The difference between D response rates in the two environments by group is shown in Figure 1 (centre). A mixed-effects logistic regression model indicated that the interaction between vowel and group is significant; thus participants succeeded at learning this complex pattern.

Experiment 3 examined whether new participants could learn a shift that was dependent on groupings of vowels rather than individual vowels. For one group of participants, the conditioning vowels were grouped by the natural classes formed by [±high] ([i, ɪ] vs. [e, ɛ]); for the other group of participants, the conditioning vowels were grouped by an unattested feature, which split those same four vowels into ([i, ɛ] vs. [ɪ, e]). The difference between D response rates in the two environments by group is shown in Figure 1 (right). Only the group whose conditioning environment formed a natural class showed a difference in D response rates between the two environments, meaning only they learned the pattern. This suggests that the natural class facilitated learning, which is the first evidence of this kind.

Figure 7: (Left) Results of Experiment 1, current *N*=28, total *N*=32; (Centre) Experiment 2, current N=30, total N=32; (Right) Experiment 3, current N=11, total N=32.



These findings show that phonological features are not only psychologically meaningful for representing contrasts and patterns, but also for the constant subconscious accommodation to complex phonetic patterns. The method itself, second-order perceptual learning, constitutes an additional contribution to this area of study: it supplies a mechanism for future work that can directly compare the salience of different phonological features in learners' minds and therefore straightforwardly test a theory's predictions about the structure of speakers' phonology.

References

Kraljic, T., Samuel, A.G. 2006. Generalization in perceptual learning for speech. Psychonomic Bulletin & Review, 13(2):262–268. / Nielsen, K. 2011. Specificity and abstractness of VOT imitation. J. of Phonetics, 39:132–142. / Norris, D., McQueen, J.M., Cutler, A. 2003. Perceptual learning in speech. Cognitive Psych., 47(2):204–238. / Onishi, K.H., Chambers, K.E., Fisher, C. 2002. Learning phonotactic constraints from brief auditory experience. Cognition, 83:B13–B23.

Realization of f0 peak alignment in Spanish by Korean L2 learners

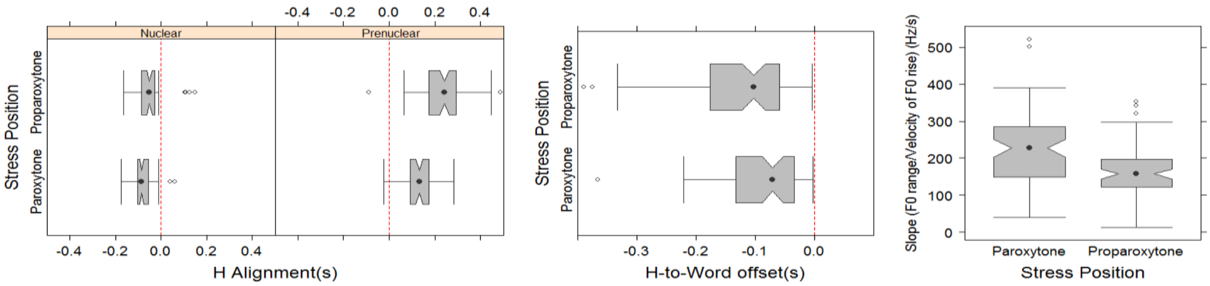
Ji Young Kim
University of Illinois at Urbana-Champaign
jkim315@illinois.edu

The objective of the present study is to examine whether Korean L2 learners of Spanish (KorL2s) produce Spanish sentences with target-like tonal alignment. In Spanish, pitch contour varies depending on the context in which a word is located. When a word is produced in nuclear position (or in isolation), the pitch accent is aligned with the stressed syllable, while in prenuclear position (of broad focus statements), a rising pitch contour is shown, in which f0 valley (L) is consistently anchored to the onset of the accented syllable, and f0 peak (H) is aligned with a syllable after the accented syllable (L+>H*) (Face & Prieto, 2007). Although there is a tendency that Spanish speakers align H with the right edge of the word, this alignment pattern is found to be variable, sensitive to factors such as stress location, syllable structure, and speech rate (Prieto & Torreira, 2007). Korean, on the other hand, is an edge prominence language with rising tonal pattern at word onset and fixed H at the end of the Accentual Phrase (AP), which is a prosodic unit containing one or more words that is higher than a phonological word and lower than an Intonational Phrase (IP) (Jun, 1993). Due to this cross-linguistic difference between Spanish and Korean in the stability of H alignment, it is hypothesized that H delay will be found in prenuclear position among KorL2s as well when they speak Spanish, but with the H consistently fixed at the right edge of the word due to influence from Korean prosody.

In total, 22 KorL2s (15F, 7M) participated in the present study, which took place in Seoul, South Korea. The subjects read out loud a list of sentences in Spanish. A total of 40 test items consisting of words of one of the two stress positions (preparoxytone vs. paroxytone) in either prenuclear or nuclear position (10 items in each condition) were used. A subset of data was taken from the productions of 10 subjects (7F, 3M) for preliminary analysis. Results show that KorL2s' H is aligned with a post-accentual syllable in prenuclear position, while in nuclear position, it is aligned with the accented syllable. This indicates that KorL2s produce H delay in prenuclear position. Moreover, H delay occurred statistically to a larger extent in proparoxytones than in paroxtyones ($t = 4.534$). Further statistical analyses on the effect of pitch accent on H-to-Word offset (or distance between H location and right edge of the word) and f0 rise slope (or f0 range / velocity of f0 rise) were conducted using liner mixed effects models with maximal random effects structure justified by the design. Results show that f0 rise slope is significantly steeper in paroxytones than in proparoxytones ($t = -3.03$). With regard to H-to-Word offset, the difference between paroxytones and proparoxytones did not reach statistical significance, indicating that H was located at the right edge of the word regardless of stress position.

Most of the tonal alignment patterns found in the preliminary data are consistent to what have been attested in Spanish native speakers (Prieto et al., 1995). Nevertheless, it is yet to be determined whether KorL2s have acquired target-like tonal alignment in Spanish. Based on auditory impression of the data, foreign accent is still detected with regard to the realization of H. This is likely to be due to stable H anchoring on the right edge of the word, which is found to be less consistent among Spanish native speakers (Prieto & Torreira, 2007). Although there have been studies that investigated various tonal alignment patterns in Spanish (Prieto et al., 1995 and Prieto & Torreira, 2007, among others), the findings of these studies cannot be directly compared to the present data, given that these studies mostly base their conclusions on a small number of subjects. Therefore, native speaker data will be additionally collected and analyzed to compare with the results found with the KorL2s, paying special attention to the degree of stability of H alignment.

Figure 8: H alignment (H - End of accented syllable) in prenuclear and nuclear position (left); H-to-word offset (middle); and f0 rising slope (f0 range/Velocity of f0 rise) in paroxytones and proparoxytones (right).



References

Face, T. & Prieto, P. (2007). Rising accents in Castilian Spanish: A revision of Sp-ToBI. *Journal of Portuguese Linguistics*, special issue on Prosody of Iberian Languages (ed. By G. Elordieta and M. Vigário), 6(1), 117–146. / Jun, S.-A. (1993). *The Phonetics and Phonology of Korean Prosody*. (Doctoral dissertation). Ohio State University. Columbus, OH. [published in 1996 by Garland Inc., New York]. / Prieto, P. & Torreira, F. (2007). The segmental anchoring hypothesis revisited: Syllable structure and speech rate effects on peak timing in Spanish. *Journal of Phonetics*, 35, 473-500. / Prieto, P.,van Santen,J., & Hirschberg, J. (1995). Tonal alignment patterns in Spanish. *Journal of Phonetics*, 23, 429–451.

The Contribution of Individual Acoustic Cues to the Perception of Focal Prominence

Tim Mahrt, Jennifer Cole

University of Illinois

tmahrt2@illinois.edu, jscole@illinois.edu

In English, speakers encode the information status of words in an utterance through prosodic prominence. For example a word may be marked for contrastive focus or as discourse new and consequently be realized with acoustic prominence. Throughout the literature there is agreement that intensity, duration, and F0, individually or in some combination, are used by speakers to mark information status, although there is disagreement over exactly which features are used and what kinds of information they encode. Breen et al (2010) in a production experiment found that the four best predictors of a word's information status included at least one measure of intensity, f0, and duration, suggesting their utility as cues to prominence. Katz and Selkirk (2011) similarly found that elements that carry contrastive focus have longer duration, greater intensity, and a greater pitch excursion than words that are discourse new. Many other works have shown that the relationship between information status and acoustic features is complex and non-obligatory (see Fery 2007 for a review). Watson (2010), in a production experiment, showed that durational prominence was due to lexical accessibility, intensity prominence was due to production effort, and F0 was used to convey pragmatic meaning such as focus. Watson's findings suggests an alternative explanation for the correlation between these three acoustic features and prominence found in earlier works, but they also raise a question about the significance of duration and intensity as cues to information status: If speakers are not using all measures of acoustic prominence to mark information status, then perhaps listeners also do not attend equally to all prominence cues in judging information status.

The present work investigates the role of duration, intensity, and F0 in the perception of broad focus and contrastive focus. The stimuli used in this study consists of ten utterances elicited under these two focus conditions and three resynthesized versions of each utterance: one resynthesized with change only in duration, one with change only in intensity, and one with change only in the F0 contour. In the resynthesis process, one acoustic parameter in a broad production was replaced with the equivalent parameter in the contrastive production (and vice versa). Resynthesis was done using PSOLA in Praat (Moulines and Charpentier, 1990; Boersma and Weenink; 2014) to replace the F0 contour and phone duration. A custom application utilizing piece-wise audio normalization was used to replace the intensity contour. These resynthesis processes do not make any assumptions about what aspect of the F0, duration and intensity patterns listeners are sensitive to. Replacement of the F0, duration or intensity pattern over the entire utterance will affect any and all aspects of the acoustic parameter that may cue the listener's perception of focus. If listeners are sensitive to any aspect of F0, duration, or intensity in distinguishing broad focus from contrastive focus, then this manipulation should change their perception of the information status of the utterance.

17 native speakers of English participated in a discrimination task administrated through a custom web-based interface developed by the first author. For each unmanipulated and resynthesized utterance, subjects had to decide whether it was an example of broad focus, contrastive focus, or neither. The results show that listeners were able to correctly distinguish the unmanipulated (source) utterances. For the manipulated (resynthesized) utterances, neither replacement of the intensity contour or phone duration resulted in a change in the perception of focus, but replacement of the F0 contour did lead to a change in perception. A few F0-manipulated stimuli showed high agreement among listeners for either no change in perception or change in perception, suggesting that listeners may have relied on auxiliary cues for these utterances. Taken together, the results from this study suggest that F0 is the most important cue used to discriminate broad focus from contrastive focus, while duration and intensity alone do not suffice to convey meaning related to focus (in line with Watson 2010), but may be used as secondary cues.

References

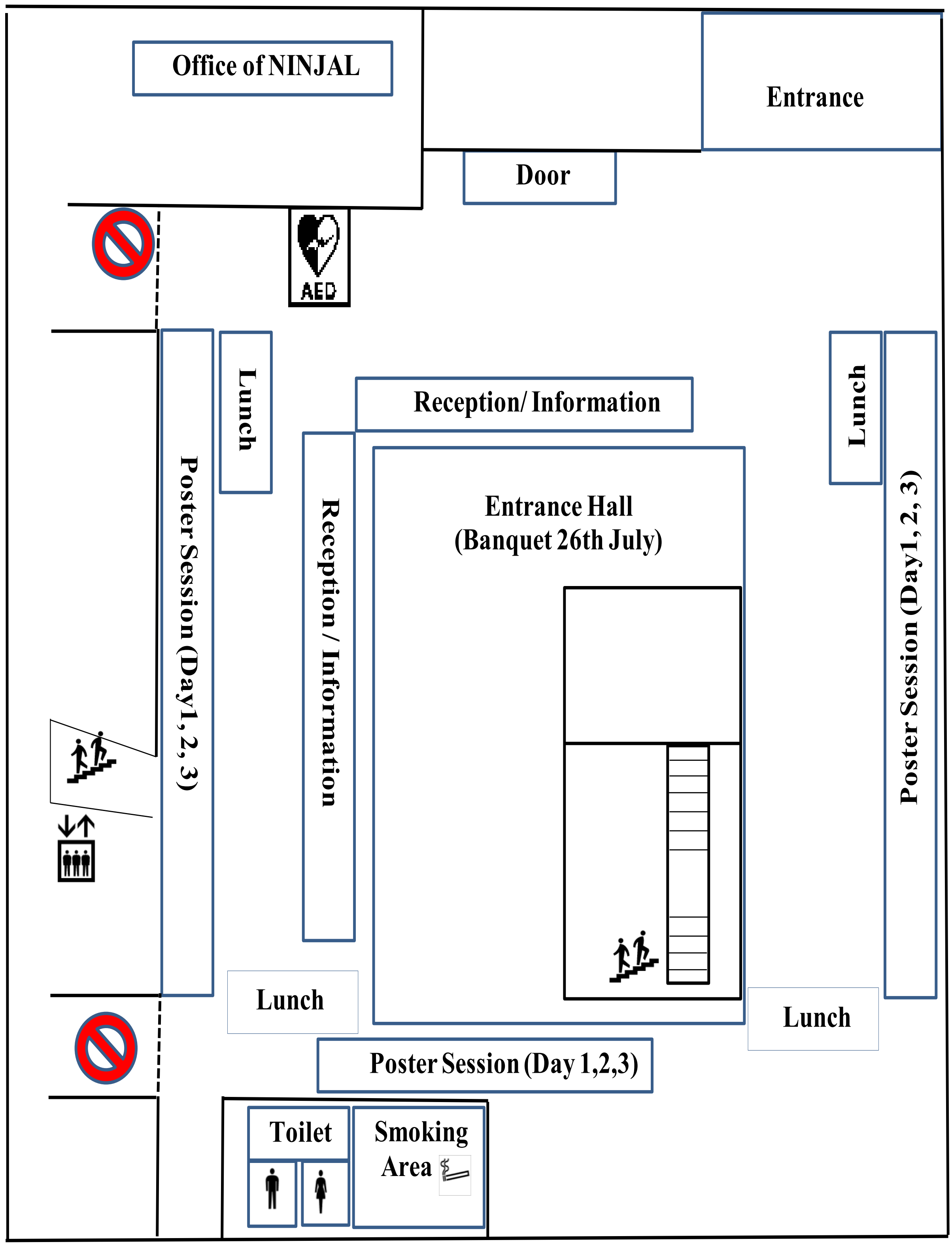
Boersma, Paul & Weenink, David (2014). Praat: doing phonetics by computer [Computer program]. Version 5.3.63, retrieved 24 January 2014 from http://www.praat.org / Breen, M., Fedorenko, E., Wagner, M., & Gibson, E. 2010. Acoustic correlates of information structure. *Language and Cognitive Processes*, *25*(7-9), 1044-1098. / Féry, C. 2008. Information structural notions and the fallacy of invariant correlates. *Acta Linguistica Hungarica*, *55*(3), 361-379. / Katz, J., & Selkirk, E. 2011. Contrastive focus vs. discourse-new: Evidence from phonetic prominence in English. Language, 87(4), 771-816. / Moulines, E., & Charpentier, F. 1990. Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech communication*, *9*(5), 453-467. / Watson, D. G. 2010. The many roads to prominence: Understanding emphasis in conversation. *Psychology of learning and motivation*, *52*, 163-183.

Author Index

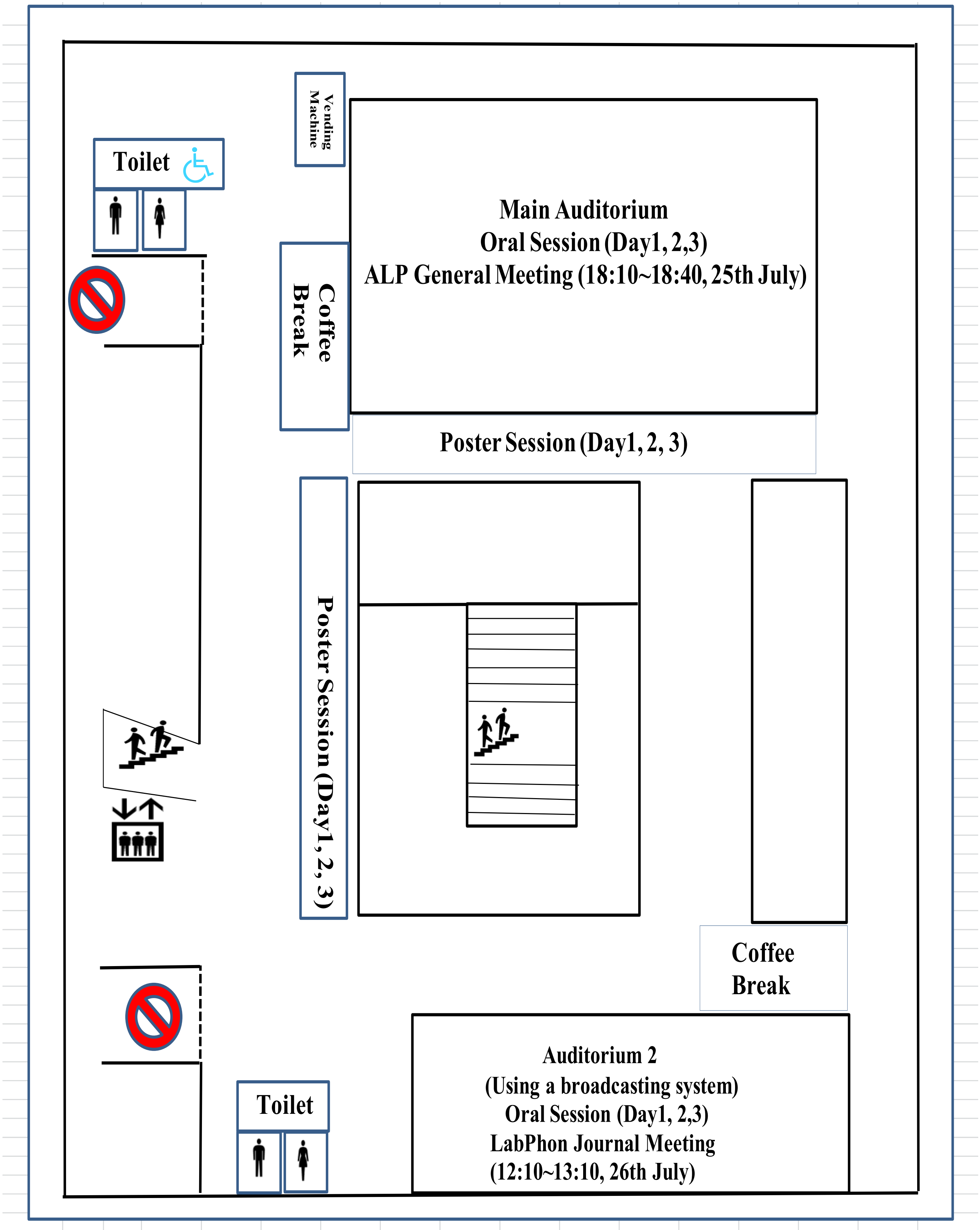
Useful Information

Map of NINJAL

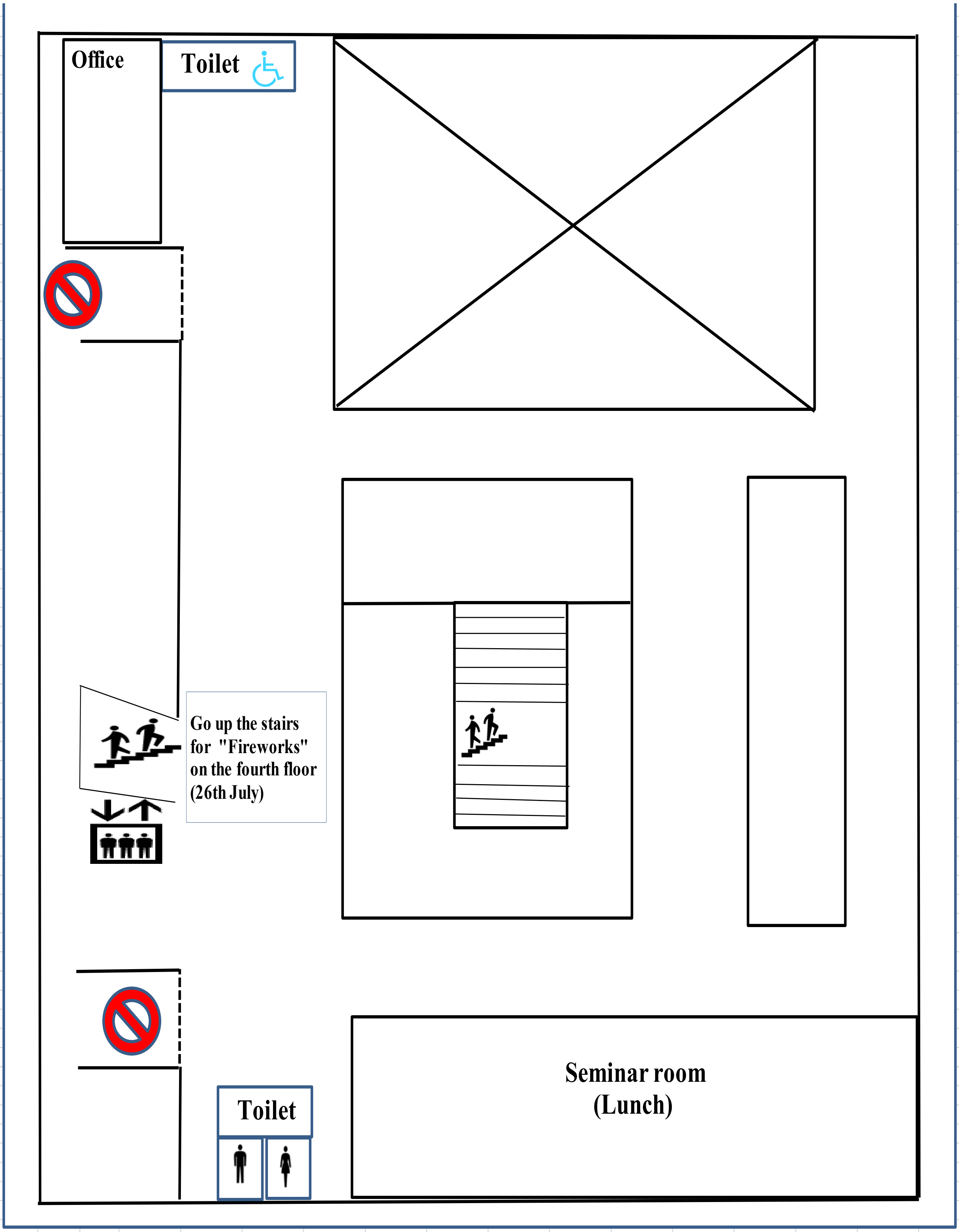
The First Floor



The Second Floor



The Third Floor



WiFi

SSID (Service Set Identifier): kokken-01
Security Key: udongenohana

Dinner Information for vegetarian

- Restaurant: マユール（Mayuuru）
  Category: Curry
  Telephone for reservation: 050-5868-2339・Telephone: 042-523-0410
  Dinner: 17:30 ~ 23:00 (Last Order 22:00)
  Budget: 2,000 Yen ~ 2,999 Yen ・Credit Card: Available (VISA, MASTER, JCB, AMEX)
  Address: 2-16-6 Akebonocho, Tachikawa, Tokyo, 190-0012
  Home-Page: http://www.mayur.jp/
- Restaurant: さんるーむ（Sanruumu）
  Category: Japanese restaurant
  Telephone: 042-540-2277・Dinner: 17:30 ~ 22:00 (Last Order 21:30)
  Budget: 1,000 Yen ~ 1,999 Yen・Credit Card: Available (VISA, MASTER, JCB, AMEX)
  Address: 7th Floor of Ganduo Tachikawa BLUG, 3-2-1, Shibasaki, Tachikawa, Tokyo, 190-0023
  Home-Page: http://www.shizenshoku-sunroom.com/
- Restaurant: 我や（Gaya）
  Category: Japanese restaurant and Alcohol
  Telephone: 042-540-9377
  Dinner: Mon ~ Sat: 17:30 ~ 24:00 (Last Order 23:00) / Sun: 17:30 ~ 23:30 (Last Order 22:30)
  Budget: 3,000 Yen ~ 3,999 Yen・Credit Card: Available (VISA, MASTER, JCB)
  Address: Kaaro BULG (カーロ ビル),1-31-1, Akebonocho, Tachikawa, Tokyo, 1900-0012
  Home-Page: http://www.gaya.co.jp/
- Restaurant: iL PePe
  Category: Buffet
  Telephone for reservation: 050-5869-1694・Telephone: 042-527-1111
  Dinner: 6:30 ~ 23:00
  Budget: 2,000 Yen ~ 2,999 Yen・Credit Card: Available (VISA, MASTER, JCB, AMEX, Diners)
  Address: 1st Floor Palace Hotel Tachikawa, 2-40-15, Akebonocho, Tachikawa, Tokyo, 1900-0012
  Home-Page: http://www.palace-t.co.jp/english/restaurant/index.html

"In Japan, there are only a few restaurants for vegetarians. You have to articulate that you do/can not eat meat or fish. And also before you eat the food, please check again if meat or fish is not in there!"

「私は菜食主義者です。肉と魚を食べられません。」"I am a vegetarian. I can/ do not eat meat and fish."

[Note]

[Note]

[Note]

[Note]

LabPhon 14

東京, 2014

LabPhon14 PROGRAM (ORAL PRESENTATIONS) (version 2014-06-24)

THURSDAY, July 24

13:30 – 17:30 Registration for LabPhon 14

14:30 – 17:30 Colloquium: D. Robert Ladd (University of Edinburgh)
Anne Cutler (University of Western Sydney)

17:30 – 19:00 Informal wine reception

FRIDAY, July 25

08:30 – 09:20 Registration

09:20 – 09:30 Welcome

SESSION 1 (Chair: Shigeto Kawahara, Keio University)

09:30 –10:20 [Invited] **Infant-directed speech as a window into the dynamic nature of phonology**
Reiko Mazuka (RIKEN Brain Science Institute)

10:20 –10:50 Comments and Discussion
Catherine T. Best (University of Western Sydney)

BREAK

10:50 –11:10

SESSION 2 (Chair: Emiko Kaneko, University of Aizu)

11:10 –11:30 **Children's imitation of coarticulatory patterns in different prosodic contexts**
Felicitas Kleber (Ludwig-Maximilians-University)
Sandra Peters (Ludwig-Maximilians-University)

11:30 –11:50 **An investigation into articulatory adaptation during acoustic mimicry of postvocalic /r/**
Eleanor Lawson (Queen Margaret University)
James M. Scobbie (Queen Margaret University)
Jane Stuart-Smith (University of Glasgow)

11:50 –12:10 **Dolls are prissy: Preschoolers' toy preferences predict medial /t/ production**
Jacqui Nokes (University of Canterbury)

LUNCH

12:10 –13:10

SESSION 3 (Chair: Mariko Kondo, Waseda University)

13:10 –13:30 **Constraints on prosodic phrasing in children's speech**
Melissa A. Redford (University of Oregon)
Zahra Foroughifar (University of Oregon)
Laura C. Dilley (Michigan State University)

13:30 –13:50 **The role of lexical age-of-acquisition on phonetic variation in natural infant-directed speech**
Georgia Zellou (University of Pennsylvania)
Rebecca Scarborough (University of Colorado)
Eric Doty (University of Pennsylvania)

13:50 –14:10 **Pre-babbling infants prefer listening to infant speech: A launch pad for the perception-production loop?**
Matthew Masapollo (McGill University)
Linda Polka (McGill University)
Lucie Ménard (Université du Québec à Montréal)

14:10-14:30 **Baby steps in perceiving articulatory foundations of phonological contrasts: infants detect audio→video congruency in native and nonnative consonants**
Catherine T. Best (University of Western Sydney)
Christian H. Kroos (University of Western Sydney)
Sophie Gates (University of Western Sydney)
Julia Irwin (University of Western Sydney)

POSTER SESSION 1 (with coffee)

14:30 –16:20

SESSION 4 (Chair: Keiichi Tajima, Hosei University)

16:30 –16:50 **Phonetic vs. phonological factors in coronal-to-dorsal perceptual assimilation**
Eleanor Chodroff (Johns Hopkins University)
Colin Wilson (Johns Hopkins University)

16:50 –17:10 **The usefulness of chaos: Lab versus non-lab speech for perceptual learning**
Elizabeth Casserly (Indiana University)
David B. Pisoni (Indiana University)

17:10 –17:30 **The effect of perceptual similarity in second language learning: Positional asymmetry in phoneme substitution**
Yu-an Lu (National Chiao Tung University)
Jiwon Hwang (Stony Brook University)

17:30 –17:50 **Sensitivity to fine acoustic detail affects comprehension of reduced speech in L2**
Ellen Aalders (Radboud University Nijmegen)
Mirjam Ernestus (Max Planck Institute Nijmegen)

17:50 –18:10 **Voicing, F0, and phonological enhancement**
James Kirby (University of Edinburgh)
D. Robert Ladd (University of Edinburgh)

ALP GENERAL MEETING
18:10-18:40

SATURDAY, July 26

SESSION 5 (Chair: Takayuki Arai, Sophia University)

09:10 –10:00 [Invited] **Labeling in the wild: Crowdsourcing versus categorical perception**
Mark Hasegawa-Johnson (University of Illinois at Urbana-Champaign)

10:00 –10:30 Comments and Discussion
Natasha Warner (University of Arizona)

BREAK

10:30 –10:50

SESSION 6 (Chair: Ian Wilson, University of Aizu)

10:50 –11:10 **The private life of stops: VOT in a real-time corpus of spontaneous Glaswegian**
Jane Stuart-Smith (University of Glasgow)
Morgan Sonderegger (McGill University)
Rachel Macdonald (University of Glasgow)
Thea Knowles (McGill University)
Tamara Rathcke (University of Kent)

11:10 –11:30 **Articulation rate and VOT in spontaneous speech**
Satsuki Nakai (University of Glasgow)
James M. Scobbie (Queen Margaret University)

11:30 –11:50 **The syllable as a prosodic unit in Japanese lexical strata: Evidence from text-setting**
Rebecca L. Starr (National University of Singapore)
Stephanie S. Shih (University of California, Berkeley)

11:50 –12:10 **Lexical diffusion of vowel length merger in Seoul Korean: A corpus-based study**
Yoonjung Kang (University of Toronto Scarborough)
Tae-Jin Yoon (Cheongju University)
Sungwoo Han (Inha University)

LUNCH / PUBLISHING IN LABORATORY PHONOLOGY: Q&A WITH THE EDITOR

12:10 –13:10

SESSION 7 (Chair: Kiyoko Yoneyama, Daito Bunka University)

13:10 –13:30 **The peril of sounding manly: A look at vocal characteristics of lawyers before the United States Supreme Court**
Alan C. L. Yu (University of Chicago)
Daniel Chen (ETH Zurich)
Katie Franich (University of Chicago)
Jacob Phillips (University of Chicago)
Betsy Pillion (University of Chicago)
Yiding Hao (University of Chicago)
Zhigang Yin (University of Chicago)

13:30 –13:50 **Stereotypes predict memory effects for voices**
Grant McGuire (University of California, Santa Cruz)
Molly Babel (University of British Columbia)

13:50 –14:10 **The role of exemplars in speech comprehension**
Annika Nijveld (Radboud University Nijmegen)
Martijn Bentum (Radboud University Nijmegen)
Louis ten Bosch (Radboud University Nijmegen)

14:10 –14:30 **Effects of phonological neighborhood density on phonetic variation: The curious case of French**
Yao Yao (Hong Kong Polytechnic University)
Christine Meunier (Aix Marseille Université)

POSTER SESSION 2 (with coffee)
14:30 –16:20

SESSION 8 (Chair: Shin-Ichiro Sano, Okayama Prefectural University)

16:30 –17:20 [Invited] **Some cognitive factors behind vowel lengthening in spontaneous Japanese: A corpus-based study**
Yasuharu Den (Chiba University)

17:20 –17:50 Comments and Discussion
Shu-Chuan Tseng (Institute of Linguistics, Academia Sinica)

BANQUET (and fireworks)
18:00 –20:00

SUNDAY, July 27

SESSION 9 (Chair: Shigeko Shinohara, Sophia University)

09:10 –10:00 [Invited] **On establishing the existence of word stress**
Carlos Gussenhoven (Radboud University Nijmegen)

10:00 –10:30 Comments and Discussion
Aditi Lahiri (University of Oxford)

BREAK

10:30 – 10:50

SESSION 10 (Chair: Shinichiro Ishihara, Goethe University Frankfurt am Main)

10:50 – 11:10 **Pointed and plateau-shaped pitch accents in North Frisian dialects**
Oliver Niebuhr (Kiel University)
Jarich Hoekstra (Kiel University)

11:10 – 11:30 **Quantity contrast in Lule Saami: A three-way system**
Nora Fangel-Gustavson (CNRS/Sorbonne Nouvelle)
Bruce Morén-Duolljá (University of Tromsø)

11:30 – 11:50 **Prominence, phrasing, and information structure in Mawng (Australian)**
Janet Fletcher (University of Melbourne)
Ruth Singer (University of Melbourne)
Deborah Loakes (University of Melbourne)

11:50 – 12:10 **No VOT perception without native VOT experience**
Rikke Bundgaard-Nielsen (La Trobe University)
Brett Baker (University of Melbourne)

LUNCH

12:10 – 13:10

POSTER SESSION 3 (with coffee)

13:10 – 15:00

SESSION 11 (Chair: Mafuyu Kitahara, Waseda University)

15:10 – 15:30 **Articulatory correlates of phonological relationships**
Kathleen Currie Hall (University of British Columbia)
Hanna Smith (University of British Columbia)
Kevin McMullin (University of British Columbia)
Noriko Yamane (University of British Columbia)
Blake Allen (University of British Columbia)
Joash Gambarage (University of British Columbia)

15:30 – 15:50 **Invariant coupling relations at the core of variable speech trajectories**
Leonardo Lancia (Max Planck Institute for Evolutionary Anthropology)

15:50 – 16:10 **Acoustic variability aids the interpretation of phonetic detail in cross-language speech production**
Lisa Davidson (New York University)
Sean Martin (New York University)
Colin Wilson (Johns Hopkins University)

16:10 – 16:30 General Discussion

16:30 – 16:35 Farewell

MONDAY, July 28 (All Day)

Satellite Workshop

Theme:	**Gestural coordination within and between speakers in first language phonological acquisition**
Organizers:	Felicitas Kleber (Ludwig-Maximilian-Universität München) Mary E. Beckman (Ohio State University)

LabPhon 14 POSTER SESSIONS (version 2014-7-8)

FRIDAY, July 25

POSTER SESSION 1 (with coffee)
14:30 –16:20

P1-1 Comparing the discourse use of prosody in adolescents with normal hearing and cochlear implants

Colleen Holt (University of Melbourne)
Ivan Yuen (Macquarie University)
David Rosson (Macquarie University)
Katherine Demuth (Macquarie University)

P1-2 Code switch production in Cantonese-English bilingual children

Donald White (The Chinese University of Hong Kong)
Peggy Mok (The Chinese University of Hong Kong)

P1-3 Modeling the acquisition of vowel normalization as cognitive manifold alignment

Andrew R. Plummer (The Ohio State University)

P1-4 Individual differences in children's prosodic focus-marking

Aoju Chen (Utrecht University/Max Planck Institute for Psycholinguistics)

P1-5 Effects of speaker language and listener language on children's stop place

Mary E. Beckman (Ohio State University, Columbus)
Benjamin Munson (University of Minnesota, Twin Cities)
Jan Edwards (University of Wisconsin, Madison)

P1-6 The acquisition of differential spectral kinematics of English sibilant fricatives

Patrick F. Reidy (The Ohio State University)

P1-7 Perception and production of phonemic vowel length in Australian English-learning 18-month-olds

Hui Chen (Macquarie University)
Nan Xu Rattanasone (Macquarie University)
Felicity Cox (Macquarie University)

P1-8 A comparison of long consonant acquisition in Arabic, Finnish, Japanese and Welsh

Ghada Khattab (Newcastle University)
Marilyn Vihman (University of York)
Jalal Al-Tamami (Newcastle University)
Satsuki Nakai (Glasgow University)
Sari Kunnari (University of Oulu)

P1-9 Do infants normalize speaker and accent variability in vowel production?

Karen E. Mulak (University of Western Sydney)
Samra Alispahic (University of Western Sydney)
Paola Escudero (University of Western Sydney)

P1-10 On lexical phonotactics and segmentability

Robert Daland (University of California, Los Angeles)
Benjamin Börschinger (Macquire University)
Abdellah Fourtassi (Laboratoire de Sciences Cognitives et Psycholinguistique, ENS/EHESS/CNRS, Paris)

P1-11 Learning sound categories with phonotactics

Masaki Noguchi (University of British Columbia)
Carla Hudson Kam (University of British Columbia)

P1-12 What kinds of units are used in speech perception?

Eva Reinisch (Ludwig Maximilian University Munich)

P1-13 Phonological and phonetic vowel reduction in two Ibero-Romance varieties

Marianna Nadeu (Penn State)

P1-14 The effect of English stress on non-native speakers' word recognition: Evidence from native speakers of lexical tone language

Shu-chen Ou (National Sun Yat-sen University)

P1-15 Evidence of early motor planning in speech production

Donald Derrick (University of Canterbury/University of Western Sydney)
Romain Fiasson (University of Western Sydney)

P1-16 Acoustic effects of predictability and gender on Japanese high vowel reduction

James Whang (New York University)

P1-17 Tapping in American English: Context matters

Adam J. Chong (University of California, Los Angeles)
Megha Sundara (University of California, Los Angeles)

P1-18 Gestural coordination in word-initial Greek clusters as revealed by kinematic, articulatory, and acoustic data

Jonathan Yip (The University of Hong Kong)

P1-19 Acoustic differentiation between coda reʈroflex and dent̪al stops in Punjabi

Qandeel Hussain (Macquarie University)
Ivan Yuen (Macquarie University)

P1-20 Stability in perceiving non-native segmental length contrasts

Asano Yuki (University of Konstanz)
Bettina Braun (University of Konstanz)

P1-21 Effect of attention on L2 speech perception: Perception of the Korean three-way stop contrast by English learners

Hyunjung Lee (University of Chicago)

P1-22 The perception of French reduced speech by native and non-native listeners

Sophie Brand (Radboud University Nijmegen)

Mirjam Ernestus (Radboud University Nijmegen/Max Planck Institute for Psycholinguistics)

P1-23 Subjective perception of affixation: A test case from Spanish

Anne Pycha (University of Wisconsin, Milwaukee)

P1-24 The psychological status of the right-branch condition and deaccentuation on *Sais-Sori* in Korean

Hyun Kyung Hwang (National Institute for Japanese Language and Linguistics)

P1-25 Perceptual relevance of Non-F0 acoustic correlates in Japanese accent

Yukiko Sugiyama (Keio University)

P1-26 The productivity and stability of competing generalizations in stress assignment

Paul Olejarczuk (University of Oregon)

P1-27 Effects of a sound change in progress on gender-marking cues in Japanese

Eun Jong Kong (Korea Aerospace University)

Kiyoko Yoneyama (Daito Bunka University)

Mary E. Beckman (Ohio State University)

P1-28 Phonological Encoding and Articulation in the Absence of Metrical Spellout

Bryan B. Holbrook (University of California, Santa Cruz)

Alan H. Kawamoto (University of California, Santa Cruz)

P1-29 Perception-production asymmetry: The case of Hai-lu Hakka tone change and tone sandhi

Chia-Hsin Yeh (Michigan State University)

Yen-Hwei Lin (Michigan State University)

P1-30 VC coarticulation in Taiwanese and its phonological consequences

Yueh-chin Chang (National Tsing Hua University)

Hsieh Feng-fan (National Tsing Hua University)

Yi-cheng Chen (National Tsing Hua University)

P1-31 Posterior cavity and aperture distance oppositions in English coronal fricatives

Simon Gonzalez (The University of Newcastle)
Mark Harvey (The University of Newcastle)
Michael Proctor (Macquarie University)

P1-32 Are acoustic cues sufficient for the identification and discrimination of Spanish approximants? Evidence from a production study

Mauricio Figueroa (University College London)
Bronwen G. Evans (University College London)

P1-33 Acoustic and lingual variation within the 4-way contrast of vowel height in French

Laurianne Georgeton (CNRS/Univ. Paris 3-Sorbonne Nouvelle)
Tanja Kocjančič Antolík (CNRS/Univ. Paris 3-Sorbonne Nouvelle)
Cécile Fougeron (CNRS/Univ. Paris 3-Sorbonne Nouvelle)

P1-34 Syntactic predictability can facilitate the recognition of casually produced words in connected speech

Malte C. Viebahn (Max Planck Institute for Psycholinguistics)
James M. McQueen (Radboud University Nijmegen)

P1-35 Perceptual attunement to coarticulation: hearing tone in vowel height

Jason A. Shaw (University of Western Sydney)
Wei-rong Chen (National Tsing Hua University
Michael D. Tyler (University of Western Sydney)
Donald Derrick (University of Western Sydney)
Michael Proctor (Macquarie University)

P1-36 Perceptual cues of Japanese /r/ Sounds: Formant transitions vs. intensity dip

Takayuki Arai (Sophia University)

P1-37 **Prosodic strengthening on initial stops in English trochaic vs. iambic words**

Jiseung Kim (University of Michigan)
Sahyang Kim (Hongik University)
Taehong Cho (Hanyang University)

SATURDAY, July 26

POSTER SESSION 2 (with coffee)
14:30 –16:20

P2-1 **Rendaku in spontaneous speech**

Shin-Ichiro Sano (Okayama Prefectural University)

P2-2 **Catalan prepalatal allophony and phonological contrast**

José I. Hualde (University of Illinois at Urbana-Champaign)
Christopher Eager (University of Illinois at Urbana-Champaign)
Marianna Nadeu (Penn State)

P2-3 **Revisiting vowel harmony in Korean sound-symbolic words: a corpus study**

Sang-Im Lee-Kim (New York University)

P2-4 **Advancing corpus-based analyses of spontaneous speech: Switch to GECO!**

Antje Schweitzer (Stuttgart University)
Natalie Lewandowski (Stuttgart University)
Grzegorz Dogil (Stuttgart University)

P2-5 **Phonetic and prosodic characteristics of disfluencies in French spontaneous speech**

George Christodoulides (University of Louvain)
Mathieu Avanzi (University Paris Diderot)

P2-6 **Multilevel modelling of initial rise in French**

Mathieu Avanzi (University Paris Diderot)
George Christodoulides (University of Louvain)

P2-7 Stress-induced sibilant variations in Taiwan Mandarin spontaneous speech

Yu-Ying Chuang (National Taiwan University)
Janice Fon (National Taiwan University)

P2-8 (t,d) deletion in everyday speech

Margaret E.L. Renwick (University of Georgia)
Rosalind Temple (University of Oxford)
Ladan Baghai-Ravary (University of Oxford)
John S. Coleman (University of Oxford)

P2-9 Using conversational corpus speech to investigate the impact of speech rate on speech perception performance over the adult life span

Xaver Koch (Radboud University Nijmegen)
Esther Janse (Radboud University Nijmegen/Donders Institute for Brain Cognition, and Behavior/Max Plank Institute for Psycholinguistics)

P2-10 Speech reduction in Czech

Alice Kolman (Radboud University Nijmegen/Christian University of Applied Sciences CHE)
Petr Pollak (Czech Technical University)

P2-11 On the tail of the Scottish vowel length rule in Glasgow

Tamara Rathcke (University of Kent)
Jane Stuart-Smith (University of Glasgow)

P2-12 Does /t/ produced as [ʔ] involve tongue tip raising? Articulatory evidence for the nature of phonological representations

Jennifer Heyward (The University of Edinburgh)
Alice Turk (The University of Edinburgh)
Christian Geng (The University of Potsdam)

P2-13 A corpus based investigation of the contrast between French rise-fall and rise via wavelet based functional mixed models

Cristel Portes (Aix-Marseille Université/CNRS/LPL)
Leonardo Lancia (Max Planck Institute for Evolutionary Anthropology)

P2-14 A new corpus of colloquial Korean and its applications

Kevin Tang (University College London)

Brent de Chene (Waseda University)

P2-15 *Phonological CorpusTools*: A free, open-source tool for phonological analysis

Scott Mackie (University of British Columbia)

Kathleen Currie Hall (University of British Columbia)

Blake Allen (University of British Columbia)

Michael Fry (University of British Columbia)

Michael McAuliffe (University of British Columbia)

P2-16 Schwa reduction in spontaneous infant-directed speech

Mybeth Lahey (Max Planck Institute for Psycholinguistics/Radboud University Nijmegen/International Max Planck Research School for Language Sciences)

Mirjam Ernestus (Radboud University Nijmegen/Max Planck Institute for Psycholinguistics)

P2-17 Uptalk in semi-spontaneous and scripted speech

Amanda Ritchart (University of California, San Diego)

Amalia Arvaniti (University of Kent)

P2-18 Variability in the phonology of child-directed speech: evidence from a new naturalistic corpus

Sam J. Green (University College London)

P2-19 The perception of phrasal prominence in conversational speech

Jennifer Cole (University of Illinois)

José I. Hualde (University of Illinois)

Tim Mahrt (University of Illinois)

Christopher Eager (University of Illinois)

Suyeon Im (University of Illinois)

P2-20 The influence of prosodic boundaries on high vowel devoicing in Japanese

Oriana Kilbourn-Ceron (McGill University)

P2-21 Production of a non-phonemic contrast by native and non-native speakers: The case of American English flap

Mafuyu Kitahara (Waseda University)
Keiichi Tajima (Hosei University)
Kiyoko Yoneyama (Daito Bunka University)

P2-22 Trajectories of phonetic variability in spontaneous speech on reality TV

Morgan Sonderegger (McGill University)

P2-23 Toward a holistic measure of reduction in spontaneous speech

Michael McAuliffe (University of British Columbia)
Molly Babel (University of British Columbia)

P2-24 Phonology Constrains the Distribution of the Particle *lah* in Singapore English

James Sneed German (Nanyang Technological University)
Laurent Prévot (Aix-Marseille Université/Laboratoire Parole et Langage)

P2-25 Word-specific and sub-phonemic representations: yod-dropping and /u/-fronting in Derby

Márton Sóskuthy (University of York)
Paul Foulkes (University of York)
Vincent Hughes (University of York)
Jennifer Hay (University of Canterbury)
Bill Haddican (CUNY-Queens College)

P2-26 Verbal feedback: positioning and acoustics of French "ouais" and "oui"

Laurent Prévot (Aix-Marseille Université/CNRS/Laboratoire Parole et Langage)
Jan Gorisch (Aix-Marseille Université/CNRS/Laboratoire Parole et Langage)

P2-27 Dissecting the consonant duration ratio

Erika Pillmeier (University of Potsdam)
Stavroula Sotiropoulou (University of Potsdam)

Adamantios Gafos (University of Potsdam)

P2-28 Derived onsets in Spanish: an experimental study of resyllabification

Patrycja Strycharczuk (CASL-Queen Margaret University)

Martin Kohlberger (Leiden University)

P2-29 Tashlhiyt syllabification: Perceptual evidence

Rachid Ridouane (CNRS/Univ. Paris 3-Sorbonne Nouvelle)

Pierre Hallé (CNRS/Univ. Paris 3-Sorbonne Nouvelle)

P2-30 Listener-specific perception of speaker-specific productions in intonation

Francesco Cangemi (University of Cologne)

Martine Grice (University of Cologne)

P2-31 Focus intonation in Turkish

Canan Ipek (University of Southern California)

Sun-Ah Jun (University of California, Los Angeles)

P2-32 Explicit and implicit gender priming in fricative perception

Benjamin Munson (University of Minnesota)

P2-33 Articulatory gestures and assimilation of mora nasal /ɴ/ in Japanese

Ai Mizoguchi (City University of New York)

Douglas H. Whalen (Haskins Laboratories)

P2-34 Aspiration metathesis in Andalusian Spanish: The role of place of articulation

Hanna Ruch (University of Zurich)

P2-35 On /tl, dl/-misperception: Language specificity and cross-linguistic tendencies

Yuriko Matsumoto-Yokoe (Sophia University)

P2-36 Vowel harmony in French: Investigating formant trajectories

Agnieszka Duniec (Université de Nantes)

Olivier Crouzet (Université de Nantes)

P2-37 Japanese nasal place/stricture assimilation: Electropalatographic evidence

Alexei Kochetov (University of Toronto)

SUNDAY, July 27

POSTER SESSION 3 (with coffee)
13:10 –15:00

P3-1 Acoustic characteristics and variation of clicks in the endangered language N|uu

Carina Marquard (Kiel University)
Oliver Niebuhr (Kiel University)
Alena Witzlack-Makarevich (Kiel University)

P3-2 Determining the representation of phonotactic restrictions with nonce words

Gillian Gallagher (New York University)

P3-3 "Chilcotin flattening" revisited: A phonetic investigation of Tsilhqut'in retraction effects

Sonya Bird (University of Victoria)

P3-4 Vowel-to-vowel coarticulation in Australian languages: Place matters

Simone Graetzer (University of Melbourne)
Janet Fletcher (University of Melbourne)
John Hajek (University of Melbourne)

P3-5 Schwa in Tashlhiyt Berber in voiceless environments

Timo B. Roettger (University of Cologne)
Rachid Ridouane (CNRS/Univ. Paris 3-Sorbonne Nouvelle)
Martine Grice (University of Cologne)

P3-6 Getting to the root of the problem: An ultrasound investigation of 'Advanced Tongue Root' in Lopit

Rosey Billington (University of Melbourne)

P3-7 Evidence of mismatch between tonal production and perception in Karbi

Amos Teo (University of Oregon)
Linda Konnerth (University of Oregon)

P3-8 Inserted vowel articulation in Scottish Gaelic: a preliminary report

Diana Archangeli (University of Arizona /University of Hong Kong)
Sam Johnston (University of Arizona)
Jae-Hyun Sung (University of Arizona)
Muriel Fisher (University of Arizona)
Michael Hammond (University of Arizona)
Andrew Carnie (University of Arizona)

P3-9 Tonal polarity in Xitsonga as a tonal artifact: an acoustic analysis

Siri Gjersøe (Humboldt University)
Seunghun J. Lee (Central Connecticut State University)

P3-10 Is longer always better? Neurolinguistic evidence for asymmetric processing of consonant duration

Sandra Kotzor (University of Oxford)
Adam Roberts (University of Oxford)
Allison Wetterlin (University of Oxford)
Aditi Lahiri (University of Oxford)

P3-11 The articulation of Kaytetye coronals

Susan Lin (University of California, Berkeley)
Mark Harvey (University of Newcastle)
Benjamin Davies (Macquarie University)
Myfany Turpin (University of Queensland)
Alison Ross (Northern Territories Department of Education)
Katherine Demuth (Macquarie University)

P3-12 Cross-dialectal vowel spaces of Greek

M. Baltazani (University of Ioannina/University of Oxford)
E. Kainada (University of Ioannina)
K. Nikolaidis (Aristotle University of Thessaloniki)
A. Sfakianaki (Aristotle University of Thessaloniki)

A. Lengeris (University of Crete/Aristotle University of Thessaloniki)
E. Tsiartsoni (Aristotle University of Thessaloniki)
D. Papazachariou (University of Patras)
M. Giakoumelou (University of Patras)

P3-13 Laryngeal and tonal contrasts in the Tai dialect of Cao Bang

Pittayawat Pittayaporn (Chulalongkorn University)
James Kirby (University of Edinburgh)

P3-14 Intra-oral pressure asindependent of duration: Stops in Bininj Gun-wok

Hywel Stoakes (University of Melbourne)

P3-15 Testing the syllabicity of inserted vowels: A word-game experiment in Pnar

Hiram Ring (Nanyang Technological Institute)
James Gruber (New Zealand Institute of Language Brain and Behavior)
Jacqui Nokes (University of Canterbury)

P3-16 Wh-question intonation patterns in the Showamura dialect of Japanese

Jason Ginsburg (Osaka Kyoiku University)
Ian Wilson (University of Aizu)
Emiko Kaneko (University of Aizu)
Naomi Ogasawara (University of Aizu)

P3-17 A study of the prosodic encoding of TOPICS as a category of information structure in Ngarinyman, a language of Australia

Candide Simard (SOAS)
Faith Chiu (University College London)
Eva Schultze-Berndt (University of Manchester)

P3-18 ERP correlates of two types of implicit knowledge of probabilistic phonotactics

Claire Moore-Cantwell (University of Massachusetts Amherst)
Lisa Sanders (University of Massachusetts Amherst)

P3-19 An EMA examination of Czech post-aveolar and palatal consonants

Phil Howson (University of Toronto)
Alexei Kochetov (University of Toronto)

P3-20 Inductive learning of long-distance dissimilation as a problem for phonology

Kevin McMullin (University of British Columbia)
Gunnar Ólafur Hansson (University of British Columbia)

P3-21 ~~Asymmetries in cross-linguistic stop/fricative perception~~ Cancelled

Youngja Nam (McGill University)
Linda Polka (McGill University)

P3-22 Acoustic disjuncture in consonant clusters and vowel epenthesis

Suyeon Yun (Massachusetts Institute of Technology)

P3-23 L2 experience can hinder perception of non-native sounds

Jeffrey Holliday (Indiana University)

P3-24 Preferences in permitted sequences: A weighted markedness constraint model

Jeremy Perkins (University of Aizu)

P3-25 Effects of linguistic structure on perceptual attention given to different speech units

Shinae Kang (University of California, Berkeley)
Keith Johnson (University of California, Berkeley)

P3-26 The perception of the tensity contrast in two German varieties

Conceição Cunha (Institute of Phonetics and Speech Processing, Munich)
Ulrich Reubold (Institute of Phonetics and Speech Processing, Munich)

P3-27 Mismatch Negativity Reveals Abstract, Monovalent Phonological Primitives

Mirjam J.I. de Jonge (University of Amsterdam)

P3-28 Phonotactic learning and its interaction with speech segmentation

Alexis Black (University of British Columbia)

Masaki Noguchi (University of British Columbia)

P3-29 Effects of a formant transition of preceding vowel off-glide on perception of Japanese geminate consonant *sokuon*

Emi Yanagisawa (Sophia University)

Takayuki Arai (Sophia University)

P3-30 An empirical assessment of the carryover bias in tonal coarticulation

Feng-fan Hsieh (National Tsing Hua University)

P3-31 Swedish focal accents and the syntax–prosody interface

Sara Myrberg (Stockholm University)

P3-32 Vowel timing in complex onsets of American English

Stavroula Sotiropoulou (University of Potsdam)

P3-33 The processing of schwa reduction in isolation and in connected speech: An electrophysiological study

Kimberley Mulder (Radboud University Nijmegen)

Linda Drijvers (Radboud University Nijmegen)

P3-34 Applying phonological knowledge to phonetic accommodation

Emily Clare (University of Toronto)

P3-35 Realization of f0 peak alignment in Spanish by Korean L2 learners

Ji Young Kim (University of Illinois at Urbana-Champaign)

P3-36 The contribution of individual acoustic cues to the perception of focal prominence

Tim Mahrt (University of Illinois)

Jennifer Cole (University of Illinois)

Information for LabPhon 14 Participants (2014.06.17)

Please visit the conference website (http://www.ninjal.ac.jp/labphon14/) from time to time as it is regularly updated. It contains many pieces of useful information including an updated program of the main conference (July 25-27) and its satellite events (July 24 and 28). It also has useful information about hotels, journey to the conference venue, tourist information, etc. Below we give some information which you might find particularly useful.

1. Hotels

   As already noted on the conference website, hotels near the conference venue are already very crowded. Most hotels in Tachikawa are already fully booked on Saturday, July 26, because of the annual Fireworks Festival in Tachikawa which attracts over 300,000 people every year. If you haven't booked your hotel yet, we recommend that you do so as soon as you can. The conference website lists more than ten hotels near the conference venue, including three big hotels that offer discount rates for LabPhon 14 participants. If you need any assistance, please contact the organizers at labphon14@ninjal.ac.jp.

2. Lunch and dinner for non-participants

   If someone accompanies you to the conference, you can book a conference dinner (Saturday, July 26) for him/her at JPY 3,000 per person (children under 12 are free). Please contact us before July 15 if you decide to bring someone with you to this banquet. In the second half of the banquet, you can enjoy the firework festival from the balcony of the building. Please note, however, that we cannot provide lunch for accompanying persons; please bring lunch for them.

3. Satellite events

   There will be a pre-conference colloquium featuring Bob Ladd and Anne Cutler on Thursday, July 24. This will be followed by registration (name tag, conference booklet, etc.) and a complimentary wine reception. There is no need to make a booking for these events. On the other hand, the satellite workshop on Monday, July 28, requires pre-registration. If you are interested in attending this workshop, please visit the following website and register for it: http://phonetik.uni-muenchen.de/labphon14-satellite/registration.html)

4. Journey to NINJAL in Tachikawa

   For those who are not familiar with Tokyo, the conference website gives details about the journey from Narita and Haneda Airports to Tachikawa and to NINJAL. Tourist information is available, too, including the links to various useful websites. In addition, complimentary guidebooks about tourists' spots in Tokyo will be available at the registration desk, in several languages including English and Japanese.

5. Wifi and PC at NINJAL

   Wifi is available free of charge at NINJAL. The password can be found in the conference booklet you will receive at the registration desk. A PC is available for oral presentations upon request. If you want to use this service, please contact us soon.

6. Restaurants near the conference venue

   There are hundreds of restaurants and shops near JR Tachikawa Station, which is 20 minutes on foot from the conference venue. Most restaurants are reasonable in price, including those on the top floor of the two big department stores near the station, Takashimaya and Isetan. At the conference, we plan to prepare a dinner map for vegetarians.

7. *The Vowel*

   The Phonetic Society of Japan is selling the last 30 copies of *The Vowel: Its nature and Structure* by Tsutomu Chiba and Masato Kajiyama. This pioneering book is out of print and costs $195 through Amazon. This book is available at the registration desk at the special conference rate of JPY1,000 (ca $10) on a first-come first-served basis.

If you have any questions or need any assistance, please contact us at labphon14@ninjal.ac.jp.

We look forward to welcoming you to Tokyo in July.

LabPhon 14 Organizers

LabPhon 14 - The 14th Conference on Laboratory Phonology

Tokyo, July 25-27, 2014

Accommodation and Travel

Acommodation around Tachikawa

- **We recommend that you make your hotel reservations well in advance (at least one month before the conference if possible). The Palace Hotel Tachikawa and the Tachikawa Grand Hotel are particularly convenient because they are both stops for the airport bus to and from Narita.**

Hotel Discount

- **The Palace Hotel Tachikawa is offering a special rate for LabPhon14 participants (Y9,700 per night for a single room). Please download this Word document for more information.**
- **Tachikawa Grand Hotel is offering special rates for LabPhon14 participants (both single and twin rooms). Please download this Word document for more information.**
- **Tachikawa Washington Hotel is offering special rates for LabPhon14 participants (single, twin and double rooms). Please download this Word document for more information.**

Links to hotels within walking distance of NINJAL

Price guide:
: most expensive
: relatively less expensive
: least expensive

- **Palace Hotel Tachikawa**

http://www.palace-t.co.jp/english/index.html
http://www.palace-t.co.jp/index.html

- **Hotel Mets Tachikawa**

http://www.jrhotelgroup.com/eng/code/codeeng160.htm
http://www.hotelmets.jp/tachikawa/

- **Tachikawa Grand Hotel**

http://tachikawa.khgrp.co.jp/

- **Tachikawa Washington Hotel**

http://tachikawa.washington-hotels.jp/
http://www.tachikawa-wh.com/

- **Toyoko Inn Tachikawa-eki Kitaguchi**

http://www.toyoko-inn.com/hotel/00245/index.html

- **Showa Hotel**

http://www.showa-hotel.jp/

- **Tachikawa Hotel**

http://www.tachikawa-h.jp/index.html

- **Business Hotel Homare**

http://hotelhomare.com/

Royal Authentic Hotel Tachikawa

http://www.authentic-hotel.co.jp/

The Crest Hotel Tachikawa

http://www.cresthotel.co.jp/tachikawa/

Hotel Rex Tachikawa

http://www.rex-tachikawa.jp/

Tachikawa Regent Hotel

http://www.tachikawa-regent-hotel.jp/

Super Hotel Tokyo-JR Tachikawa Kitaguchi

http://www.superhotel.co.jp/s_hotels/tachikawakita/tachikawakita.html

Tourist Information

Japan

http://www.att-japan.net/
http://www.japan-guide.com/
http://travel.rakuten.com/

Tokyo

http://www.gotokyo.org/en/index.html
http://guide.enjoytokyo.jp/en/

Tachikawa

http://www.tbt.gr.jp/english/index.html

Yokohama

http://www.welcome.city.yokohama.jp/eng/travel/

Kamakura

http://en.kamakura-info.jp
http://guide.city.kamakura.kanagawa.jp/

Hakone

http://www.hakone.or.jp/english/index.html

Nikko

http://www.nikko.or.jp/index.html
http://nikko.4-seasons.jp/

Kyoto

http://www.kyoto-magonote.jp/en/
http://www.kyoto.travel/

Osaka

http://www.osaka-info.jp/en/

Back to home Back Top

Sitemap

10-2 Midori-cho, Tachikawa City, Tokyo, Japan 190-8561

Walking
Tachikawa Station to NINJAL

These directions assume that your starting point is inside Tachikawa Station near the ticket gates that train passengers pass through.

[1] Exit the Station

Begin by finding the **north exit**. The ticket gates are one level above ground level, and when you exit the station, you will be on an elevated plaza. Turn left and walk west.

looking north from the north exit of Tachikawa Station

[2] Pass the Monorail Entrance

As you walk west across the elevated plaza, you will see the Mizuho Bank straight ahead and the monorail station on your right. Continue past the monorail station escalator and keep walking west as far as you can (without going into a building). Then turn right (north).

looking west across the plaza


monorail station escalator

[3] Follow the Elevated Walkway

As you walk north you will see Mister Donut straight ahead. Just before Mister Donut there is a jog in the walkway, and the surface changes from artificial wood to orange asphalt.

elevated walkway

[4] Go Down the Stairs

When you get to the end of the elevated walkway, go down the stairs to the landing, turn right, and then continue down the stairs to street level.

stairway to street level

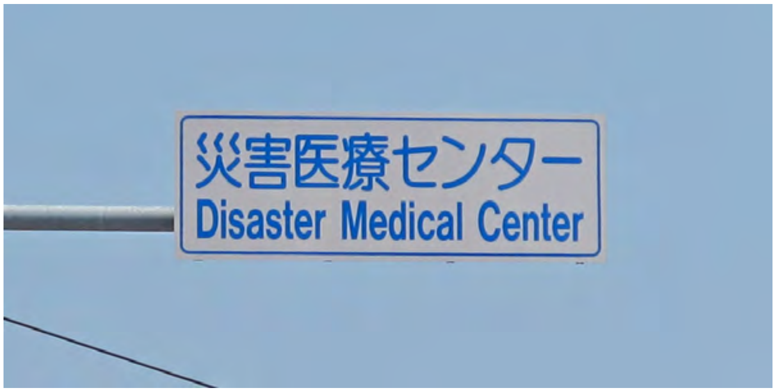
[5] Follow the Sidewalk to Shōwa Kinen Park

Turn left at the bottom of the stairs and walk north on the sidewalk. There will be three major intersections (with traffic lights) before you get to NINJAL. The first is by the entrance to Shōwa Kinen Park.

Shōwa Kinen Park intersection

[6] Pass the National Hospital

The second major intersection is by the entrance to the National Hospital, known as the Disaster Medical Center. IKEA is across the street (to the east). Just keep walking north.



Disaster Medical Center intersection

[7] Cross the Unmarked Intersection

The next intersection, after the National Hospital, is unmarked and has no traffic light.

[8] Pass the Local Autonomy College

The last major intersection before NINJAL is near the National Autonomy College (Jichi Daigakkō). Confusingly, the sign above the intersection says “Jichidaigaku” instead.

National Autonomy College intersection

[9] Turn Left at the NINJAL Bus Stop

Shortly after the National Autonomy College, you will see the bus stop in front of NINJAL.

NINJAL bus stop

You made it!

Riding the Bus
Tachikawa Station to NINJAL

These directions assume that your starting point is inside Tachikawa Station near the ticket gates that train passengers pass through.

[1] Exit the Station

Begin by finding the **north exit**. The ticket gates are one level above ground level, and when you exit the station, you will be on an elevated plaza. Turn left and walk west.

looking north from the north exit of Tachikawa Station

[2] Find the Stairway for Bus Stop Number 2

You will see an overhead sign with an arrow pointing to the stairway for bus stops 1, 2, and 3. You will be facing east as you go down, and stop number 2 is the nearest one to the foot of the stairs.

stairway for bus stop #2

[3] Get in Line at Bus Stop Number 2

All buses that leave from stop number 2 stop at NINJAL. If the bus is already there, just get on. If not, there will probably be a line of people waiting, so just go to the end of the line. The guy wearing a baseball cap in the picture below is first in line.

[4] Take a Boarding Ticket

Board the bus through the rear door and pull a boarding ticket (*seiriken*) from the machine just inside the door (shown in the picture at the left). You will need to hold onto the ticket until you get off. (Your bus fare depends on how far you travel, and your boarding ticket has a number printed on it to show where you got on.)

[5] Prepare Exact Change

Make sure have exact change. The fare is ¥190. If necessary, you can break a ¥1,000-yen note or a ¥500 or ¥100 coin in the machine next to the driver (shown in the picture on the right). The slot for paper money is on the front, where the upward-pointing red arrow is. The slot for coins is on top. The machine does not subtract your bus fare from the amount you put in; it simply gives you coins in small enough denominations so that you can pay your fare with exact change. If you need to do this, try to do it as soon as you get on the bus so that you will not delay the bus by making change when you need to get off. You cannot actually pay your fare until you get off.

Almost everyone who lives in Tokyo carries an IC (integrated circuit) card for paying bus and train fares, so you may be the only person on your bus who actually takes a boarding ticket and pays with cash.

[6] Listen for Your Stop

There are recorded announcements and visual displays that tell passengers what the next stop is, but these are in Japanese only. NINJAL is at the fourth stop on the route. The recorded announcement will say *Jichi daigakkō*, *Kokuritu kokugo kenkyūjo*, and the visual display will read 自治大学校･国立国語研究所. The preceding three stops are:

(1) *Akebono-chō* 曙町
(2) *Saigai iryō sentā higashi* 災害医療センター東
(3) *Midori-chō* 緑町

[7] Request Your Stop

There will be a stop-request button (like the one in the picture at the right) within reach no matter where you are sitting or standing on the bus, and the driver will not stop unless someone pushes one of these buttons. Pushing the button triggers a chiming sound, followed by an announcement that the bus will stop at the next stop. If you happen to be the only person on your bus who wants to get off at NINJAL, you will have to push the button yourself. If you miss the stop and see the NINJAL building out the windows on the left side of the bus, push one of the buttons right away and get off at the next stop. It's only a little farther from the front door to NINJAL.

[8] Pay Your Fare

Drop both your boarding ticket and ¥190 in coins into the fare box on top of the machine next to the driver (shown in the picture at the left) and exit through the front door. Watch out for reckless bicyclists as you step onto the sidewalk.

[7] Returning to Tachikawa Station

If you to take the bus from NINJAL back to Tachikawa Station, you will have to walk north from the NINJAL building to the crosswalk at the traffic signal. Cross to the opposite side of the big street that runs in front of the NINJAL building and go to the bus shelter on your right. You will need to take a boarding ticket and pay your fare with exact change, but you will not have to worry about pushing the stop-request button, since Tachikawa Station is the last stop.

Please be advised that the buses going to Tachikawa Station are often very crowded at rush hour, especially if it happens to be raining, and they are often several minutes behind schedule.

You are likely to see buses pass by without stopping. These are on different routes and are not supposed to stop at NINJAL.

bus shelter across the street from NINJAL

Riding the Monorail
Tachikawa Station to NINJAL

These directions assume that your starting point is inside Tachikawa Station near the ticket gates that train passengers pass through.

[1] Exit the Station

Begin by finding the **north exit**. The ticket gates are one level above ground level, and when you exit the station, you will be on an elevated plaza. Turn left and walk west.

looking north from the north exit of Tachikawa Station

looking west across the plaza

[2] Walk to the Monorail Escalator

As you walk west across the elevated plaza, you will see the Mizuho Bank straight ahead and the monorail station on your right.

[3] Take the Escalator up to Tachikawa-Kita Station

There is an escalator (and also a stairway) going up into the monorail station (Tachikawa-Kita Station). Tokyo escalator etiquette is to keep to the left side if you stand still. People who are in a hurry walk on the right side of moving escalators.

monorail station escalator

[4] Buy at Ticket

Look for the ticket vending machines on your left as you walk forward, and buy a ¥100 ticket. If you don't have exact change, put in a ¥500 coin or a bill. When you put your money in, the screen will light up and give you several choices for your ticket price. Push the "100" button. (It is not an actual button, of course; it is just a touch sensitive area on the screen.) Your ticket (and your change) will come out.

[5] Insert Your Ticket into the Ticket Gate

Put your ticket into the slot on the ticket gate. It will come out the other side as you move through the gate. Don't forget to take the ticket back and hold onto it. You will need it when you exit at your destination.

[6] Take the Escalator up to Track 1

The northbound trains stop at Track 1. Take the escalator up to the platform.

[7] Take Any Northbound Train to Takamatsu

Your destination is Takamatsu, the first stop north of Tachikawa-Kita. The trains run every 10 minutes most of the day. All northbound trains go at least as far as Takamatsu, so board the first train that comes.

[8] Get off at Takamatsu

Get off the train at Takamatsu Station and take the stairs down to the ticket gate level. Put your ticket into the slot on the ticket gate. The gate will eat your ticket and say thank you. Take the stairway on the west side of the station down to street level.

[9] Walk to NINJAL

Follow the map at the right and walk to NINJAL.

Taxi Map

立川市緑町 10-2　　国立国語研究所