Victoria
UNIVERSITY OF WELLINGTON
Te Whare Wānanga o te Ūpoko o te Ika a Māui

Programme Abstracts Links, etc.

WELCOME TO THE

Laboratory Phonology 11 Conference

- **Home**
- **Abstracts**
  - View pdf versions of abstracts here
- **Programme**
  - Full programme
  - Poster sessions
- **Flyers & posters, Links, etc.**

where is Wellington? where is New Zealand? click on the map for more information

Additional sponsors:

UC UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

ASSTA
The Australasian Speech Science & Technology Association Incorporated

Australasian Speech Science & Technology Association

International Conference Fund
(Funded by the NZ Government and administered by the Royal Society of NZ)

Wellington, New Zealand
30 June - 2 July 2008

LabPhon 11 was held at Victoria University of Wellington, New Zealand, 30 June – 2 July 2008, and was followed at the same venue by the 5th International Gender and Language Association conference (IGALA5), 3-5 July 2008

The overall theme of the LabPhon 11 conference was ***Phonetic detail in the lexicon***, with the sub-themes shown below. Invited plenary speakers and commentators are listed against these sub-themes.

- **Accessing the lexicon**
  Invited speaker:**Keith Johnson, UC Berkeley**
  Commentator: **Marcus Taft, University of New South Wales**
  - How stable are lexical representations?
  - What encoding takes place between the signal and the lexicon?
  - To what degree does literacy/orthography influence lexical representation and access
- **Social information in the lexicon**:
  Invited speaker: **Paul Foulkes, University of York**
  Commentator: **Ben Munson, University of Minnesota**
  - Is phonetic information in the lexicon accompanied by social information?
  - How do social expectations about a speaker affect speech perception?
  - Is speaker-specific detail stored in the lexicon?
- **Phonetic cues to lexical structure**
  Invited speaker: **Alice Turk, University of Edinburgh**
  Commentator: **Laurie Bauer, Victoria University of Wellington**
  - To what degree does phonetic detail reflect morphological structure?
  - How are different types of boundaries expressed phonetically?
  - Is there phonetic evidence of lexical entries greater than the word?
- **Generalising over the lexicon**
  Invited speaker: **Anne Cutler, Max Planck Institute for Psycholinguistics & MARCS laboratory**
  Commentator: **Karen Croot, University of Sydney**
  - Is there evidence for or against the existence of abstract phonological categories?
  - What types of statistical generalisations emerge from the lexicon?
  - What is the nature of the relationship between the content of the lexicon and the shape of the phonological grammar?

Updated: 13 March, 2012

Victoria UNIVERSITY OF WELLINGTON Te Whare Wānanga o te Ūpoko o te Ika a Māui

Programme

All take place in RH-LT1 and/or Mezzanine Foyer, Rutherford House (see "venue")

Monday June 30			
	8:00-	Registration / conference pack pick-up	
	8:45-9:00	Prof Neil Quigley DVC(Research)	Welcome
Accessing the lexicon	9:00-9:50	Keith Johnson	Invited Plenary: The linguistic basis of phonetic coherence
	9:50-10:20	Jane Stuart-Smith, Rachel Smith & Sophie Holmes	Phonological learning based on interactive and mediated speech
	10:20-10:50	Michael Key	Interactive and autonomous modes of speech perception: Phonological knowledge and discrimination in English and French listeners
	10:50-11:20	morning tea	
	11:20-11:50	Baris Kabak, Kazumi Maniwa & Nina Kazanina	Listeners use vowel harmony and word-final stress to spot nonsense words: A study of Turkish and French
	11:50-12:30	Marcus Taft	Comments
	12:30-1:30	lunch	
Phonetic cues to lexical structure	1:30-2:20	Alice Turk	Invited Plenary: Prosodic constituency signals relative predictability
	2:20-2:50	Azra Ali & Michael Ingleby	Gradience in morphological binding: Evidence from perception of audiovisually incongruent speech
	2:50-3:20	Shari Speer & Lei Xu	Processing lexical tone in third-tone sandhi
	3:20-3:50	afternoon tea	
	3:50-4:20	Travis Wade & Bernd Möbius	Detailed phonetic memory for multi-word and part-word sequences
	4:20-5:00	Laurie Bauer	Comments
	5:00-6:30	poster session 1	
		reception	

Tuesday July 1			
	9:00-9:50	Paul Foulkes	Invited Plenary: Exploring social-indexical knowledge: a long past but a

Social information in the lexicon			short history
	9:50-10:20	E. Allyn Smith, Kathleen Currie Hall & Benjamin Munson	Bringing semantics to sociophonetics: Social variables and secondary entailments
	10:20-10:50	Katie Drager	Sensitivity to Grammatical and Sociophonetic Variability in Perception
	10:50-11:20	morning tea	
	11:20-11:50	Cynthia G. Clopper, Janet B. Pierrehumbert & Terrin N. Tamati	Lexical Bias in Cross-Dialect Word Recognition in Noise
	11:50-12:20	Sara Mack	A sociophonetic analysis of perception of sexual orientation in Puerto Rican Spanish
	12:20-1:00	Ben Munson	Comments
	1:00-2:00	lunch	
non-thematic	2:00-2:30	Daylen Riggs & Dani Byrd	The Scope of Phrasal Lengthening in Articulation: Prosody and Prominence
	2:30-3:00	Jennifer Cole, Yoonsook Mo & Mark Hasegawa-Johnson	Frequency and repetition effects outweigh phonetic detail in prominence perception
	3:00-3:30	afternoon tea	
	3:30-4:00	Felicity Cox & Sallyanne Palethorpe	Nasalisation of /æ/ and sound change in Australian English
	4:00-4:30	Abigail Cohn & Anastasia Riehl	The Internal Structure of Nasal-Stop Sequences: Evidence from Austronesian
	4:30-6:00	poster session 2	
	7:00-9:30	Dinner: The Brewery Bar & Restaurant Cnr Cable & Taranaki Sts	
Wednesday July 2			
Generalising over the lexicon	9:00-9:50	Anne Cutler	Invited Plenary: Efficiency in the Lexicon: From Lexical Statistics to Abstract Generalisations
	9:50-10:20	Mary Beckman & Jan Edwards	Generalizing over Lexicons to Predict Consonant Mastery
	10:20-10:50	Stefan Frisch, Maria Brea-Spahn & Carla Orellana	Metalinguistic judgments of phonotactics by bilinguals
	10:50-11:20	morning tea	
	11:20-11:50	Vsevolod Kapatsinski	Rule Reliability and Productivity: Velar Palatalization in Russian and Artificial Grammar
	11:50-12:20	Mary Ann Walter	Vowel Cooccurrence in the Lexicon: An Anti-OCP Effect?
	12:20-1:00	Karen Croot	Comments
	1:00-2:00	lunch	
	2:00-2:30	Attila Andics & James M. McQueen	Plasticity in voice category learning

non-thematic	2:30-3:00	Molly Babel & Keith Johnson	Accessing psycho-acoustic perception and language-specific perception with speech sounds
	3:00-3:30	afternoon tea	
	3:30-4:00	Louis Goldstein, Hosung Nam, Manisha Kulshreshtha, Leslie Root & Catherine T. Best	Distribution of tongue tip articulations in Hindi versus English and the acquisition of stop place categories
	4:00-4:30	general discussion	
	5:30-7:00	Mayor's reception: Wellington Town Hall (for LabPhon11 and IGALA5)	

Conference Abstracts

By clicking on the link below you will be able to access the pdf file "LabPhon11.pdf" which containts the entire set of 85 two-page abstracts for the conference. (Note: invited speakers were not required to submit an abstract.) The "bookmarks" tab in the pdf file will give you access to an alphabetical index, by author lastname for all authors (i.e. including all authors of multiple authored papers). The entire file is quite large, and you may want to download it for your own future reference. The individual authors are listed alphabetically below, with their paper titles.

LabPhon11.pdf

Alphabetical list by author surname (all authors included)

- Frans ADRIAANS
  - Learning Phonotactic Generalizations from Continuous Speech: A Computational Study
- Eleonora ALBANO
  - Morphological And Phonetic Factors In Brazilian Portuguese Lexical Phonotactics
- Azra ALI
  - Gradience in Morphological Binding: Evidence from Perception of Audiovisually Incongruent Speech
  - Experiments on Fine Phonetic Detail of Arabic Syllable Structure
- Jalal AL-TAMIMI
  - Phonetic Cues to Gemination in Lebanese Arabic
- Attila ANDICS
  - Plasticity in voice category learning
- Arto ANTTILA
  - The Gradient Phonotactics of English CVC Syllables
- Mark ANTONIOU
  - Greek-Australian bilinguals match the VOTs of Greek and Australian English native speakers depending on language context

- Molly BABEL
  - Accessing psycho-acoustic perception and language-specific perception with speech sounds
- William BADECKER
  - Front/Back Asymmetries in Height Harmony
- Winifred BAUER
  - Ka conversion – the changing sound and rhythm of Māori?
- Stefan BAUMANN
  - Articulatory Marking of Focus in German
- Johannes BECKER

    - Articulatory Marking of Focus in German
- Mary BECKMAN
    - Generalizing over Lexicons to Predict Consonant Mastery
    - Adult Acoustics and Developmental Patterns for Gender-marked Phonetic Variants in Mandarin Fricatives and Japanese Stops
- Catherine BEST
    - Greek-Australian bilinguals match the VOTs of Greek and Australian English native speakers depending on language context
    - Distribution of tongue tip articulations in Hindi versus English and the acquisition of stop place categories
- Natalie BOLL-AVETISYAN
    - Identity Avoidance Between Non-Adjacent Consonants in Artificial Language Segmentation
- Ivana BRASILEIRO
    - The Perception of a Vowel Contrast by Bilingual Children
- Maria BREA-SPAHN
    - Metalinguistic judgments of phonotactics by bilinguals
- Susanne BROUWER
    - How Acoustically Reduced Forms Activate the Lexicon: Evidence from Eye-tracking
- Jason BROWN
    - Gradient phonotactics in the Gitksan lexicon
- Marc BRYSBAERT
    - Can auditory distractors disrupt speech execution?
- Dani BYRD
    - The Scope of Phrasal Lengthening in Articulation: Prosody and Prominence

- Basilio CALDERONE
    - Recovering morphology from local phonotactic constraints
- Chiara CELATA
    - Recovering morphology from local phonotactic constraints
- Della CHAMBLESS
    - The independence of auditory and categorical effects on speech perception
- Cynthia CLOPPER
    - Lexical Bias in Cross-Dialect Word Recognition in Noise
- Abigail COHN
    - The Internal Structure of Nasal-Stop Sequences: Evidence from Austronesian
- Jennifer COLE
    - Frequency and repetition effects outweigh phonetic detail in prominence perception
- Felicity COX
    - Nasalization of /æ/ and sound change in Australian English

- Mariapaola D'IMPERIO
    - Phonetic variation and phonological phrasing: does the Accentual Phrase exist in Italian?
- Olga DMITRIEVA
    - The Gradient Phonotactics of English CVC Syllables
- Gerard DOCHERTY
    - Perceptual evaluation of sociophonetic variability: how do listeners learn?
    - An investigation of cross-language differences in pitch range for speakers of English and German
- Grzegorz DOGIL
    - Subglottal Resonances and Vowel Formant Variability: A Case

  - Study of High German Monophthongs and Swabian Diphthongs
- Katie DRAGER
  - Sensitivity to Grammatical and Sociophonetic Variability in Perception

- Jan EDWARDS
  - Generalizing over Lexicons to Predict Consonant Mastery
  - Adult Acoustics and Developmental Patterns for Gender-marked Phonetic Variants in Mandarin Fricatives and Japanese Stops
- Mirjam ERNESTUS
  - How Acoustically Reduced Forms Activate the Lexicon: Evidence from Eye-tracking
- Carsten EULITZ
  - Experimental approaches to lexical representations of tone in Swedish: What is stored and how does it affect lexical word access?

- Verena FELDER
  - : What is stored and how does it affect lexical word access?
- Sara FINLEY
  - Front/Back Asymmetries in Height Harmony
- Paul FOULKES
  - Exploring social-indexical knowledge: a long past but a short history
  - Perceptual evaluation of sociophonetic variability: how do listeners learn?
- Carol FOWLER
  - Compensation for Coarticulation may reflect gestural perception: Evidence from a critical examination of the effects of non-speech contexts on speech categorization
- Stefan FRISCH
  - Metalinguistic judgments of phonotactics by bilinguals

- Adamantios GAFOS
  - Obligatory release and stiffness modulation in Moroccan Arabic
- Chip GERFEN
  - The time course of stress and syllable-level encoding: Evidence from a delayed naming task in Spanish
- Andy GIBSON
  - Perception of Sung and Spoken Vowels in New Zealand English
- Louis GOLDSTEIN
  - Distribution of tongue tip articulations in Hindi versus English and the acquisition of stop place categories
- Izabelle GRENON
  - That elusive rhythm: Pros and cons of rhythm metrics
- Martine GRICE
  - Articulatory Marking of Focus in German
- Nicolas GUTIERREZ PALMA
  - The time course of stress and syllable-level encoding: Evidence from a delayed naming task in Spanish
- Gregory GUY
  - Phonological, lexical, and frequency factors in coronal stop deletion in early New Zealand English

- Kathleen Currie HALL
  - Phonological Relationships: Beyond Contrast and Allophony
  - Bringing Semantics to Sociophonetics: Social Variables and Secondary Entailments

- Gunnar HANSSON
  - Gradient phonotactics in the Gitksan lexicon
- Ray HARLOW
  - Ka conversion – the changing sound and rhythm of Māori?
- Mark HASEGAWA-JOHNSON
  - Frequency and repetition effects outweigh phonetic detail in prominence perception
- Sarah HAWKINS
  - Syllabifications of the /st/ cluster and vowel-to-vowel coarticulation in English
- Jen HAY
  - Phonological, lexical, and frequency factors in coronal stop deletion in early New Zealand English
- Anne HERMES
  - Articulatory Marking of Focus in German
- Ivan HERREROS
  - Recovering morphology from local phonotactic constraints
- Sophie HOLMES
  - Phonological learning based on interactive and mediated speech
- Phil HOOLE
  - Obligatory release and stiffness modulation in Moroccan Arabic
- John HOUDE
  - Compensation ?=? Mental Representation
- Fang HU
  - When Ongoing Historical Sound Change Meets with Social Factors
- Hsiao-Han HUANG
  - Lexical Tonal Space and Sandhi Rule

- Michael INGLEBY
  - Experiments on Fine Phonetic Detail of Arabic Syllable Structure
  - Gradience in Morphological Binding: Evidence from Perception of Audiovisually Incongruent Speech

- Keith JOHNSON
  - Accessing psycho-acoustic perception and language-specific perception with speech sounds
- Elisabet JOENSSON-STEINER
  - Experimental approaches to lexical representations of tone in Swedish: What is stored and how does it affect lexical word access?

- Baris KABAK
  - Listeners Use Vowel Harmony and Word-Final Stress to Spot Nonsense Words: A Study of Turkish and French
- Rene KAGER
  - Identity Avoidance Between Non-Adjacent Consonants in Artificial Language Segmentation
- Evia KAINADA
  - Influence of Prosodic Phrasing on Stop Voicing and Vowel Hiatus Resolution in Modern Greek
- Eden KAISER
  - Social Selectivity in Adults' Novel Sound Learning
- Vsevolod KAPATSINSKI
  - The role of phonetic detail in associating phonological units
  - Rule Reliability and Productivity: Velar Palatalization in Russian and Artificial Grammar
- Shira KATSEFF
  - Compensation ?=? Mental Representation

- Shigeto KAWAHARA
  - The independence of auditory and categorical effects on speech perception
- Nina KAZANINA
  - Listeners Use Vowel Harmony and Word-Final Stress to Spot Nonsense Words: A Study of Turkish and French
- Peter KEEGAN
  - Ka conversion – the changing sound and rhythm of Māori?
- Michael KEY
  - Interactive and autonomous modes of speech perception: Phonological knowledge and discrimination in English and French listeners
  - The independence of auditory and categorical effects on speech perception
- Ghada KHATTAB
  - Phonetic Cues to Gemination in Lebanese Arabic
- Jeanette KING
  - Ka conversion – the changing sound and rhythm of Māori?
- John KINGSTON
  - The independence of auditory and categorical effects on speech perception
- Naoko KINOSHITA
  - The acquisition of lexical rhythm and duration by Japanese second language learners
- Eunjong KONG
  - Adult Acoustics and Developmental Patterns for Gender-marked Phonetic Variants in Mandarin Fricatives and Japanese Stops
- Christian KROOS
  - Greek-Australian bilinguals match the VOTs of Greek and Australian English native speakers depending on language context
- Manisha KULTHRESHTHA
  - Distribution of tongue tip articulations in Hindi versus English and the acquisition of stop place categories

- Aditi LAHIRI
  - An investigation of unaspirated dental and retroflex stops in Bangla
  - Experimental approaches to lexical representations of tone in Swedish: What is stored and how does it affect lexical word access?
  - Harmful Reduction?
- Christian LANGSTROF
  - Perceptual evaluation of sociophonetic variability: how do listeners learn?
- Eleanor LAWSON
  - Having your cake and eating it: An articulatory perspective on the individual's place in systems of variation and change
- Chia-Chien LEE
  - Lexical Tonal Space and Sandhi Rule
- Yunwei LEE
  - Phonetic Details and Lexical Idiosyncrasies in Vowel Categorization in Taiwanese
- Fangfang LI
  - Adult Acoustics and Developmental Patterns for Gender-marked Phonetic Variants in Mandarin Fricatives and Japanese Stops
- Pi-Chiang LI
  - Lexical Tonal Space and Sandhi Rule
- Deborah LOAKES
  - Phonetic Evidence for Neutralisation of Prelateral /ae/ and /e/ in

Melbourne English

- Steven LULICH
    - Subglottal Resonances and Vowel Formant Variability: A Case Study of High German Monophthongs and Swabian Diphthongs
- James McQUEEN
    - Plasticity in voice category learning
- Sara MACK
    - A sociophonetic analysis of perception of sexual orientation in Puerto Rican Spanish
- Margaret MACLAGAN
    - Ka conversion – the changing sound and rhythm of Māori?
- Ian MADDIESON
    - Segmentation, not just similarity: alliteration as a natural experiment
- Andreas MADSACK
    - Subglottal Resonances and Vowel Formant Variability: A Case Study of High German Monophthongs and Swabian Diphthongs
- Katalin MADY
    - Heads you lose: prosodic structure and timing in Hungarian
- James MAGNUSON
    - Compensation for Coarticulation may reflect gestural perception: Evidence from a critical examination of the effects of non-speech contexts on speech categorization
- Kazumi MANIWA
    - Listeners Use Vowel Harmony and Word-Final Stress to Spot Nonsense Words: A Study of Turkish and French
- Sven MATTYS
    - That elusive rhythm: Pros and cons of rhythm metrics
- Ineke MENNEN
    - An investigation of cross-language differences in pitch range for speakers of English and German
- Holger MITTERER
    - How Acoustically Reduced Forms Activate the Lexicon: Evidence from Eye-tracking
- Yoonsook MO
    - Frequency and repetition effects outweigh phonetic detail in prominence perception
- Bernd MOEBIUS
    - Detailed Phonetic Memory for Multi-Word and Part-Word Sequences
- Peggy MOK
    - Syllabifications of the /st/ cluster and vowel-to-vowel coarticulation in English
- Christine MOOSHAMMER
    - Intrinsic Pitch is not a universal phenomenon: Evidence from Romance languages
- Rebecca MORLEY
    - Integrating Sub-phonemic Cues: learning across a morpheme boundary
- Doris MUECKE
    - Articulatory Marking of Focus in German
- Benjamin MUNSON
    - Social Selectivity in Adults' Novel Sound Learning
    - Bringing Semantics to Sociophonetics: Social Variables and Secondary Entailments
- Naomi NAGY
    - Perceptual Frequency and Formant Frequency in R speech

- Hosung NAM
  - Distribution of tongue tip articulations in Hindi versus English and the acquisition of stop place categories
- Partha NIYOGI
  - When does variation lead to change? A dynamical systems model of a stress shift in English

- Naomi OGASAWARA
  - Does what listeners hear affect what they say? Evidence from shadowing experiment
- John OHALA
  - Voicing control and nasalization
- Carla ORELLANA
  - Metalinguistic judgments of phonotactics by bilinguals

- Sallyanne PALETHORPE
  - Nasalization of /æ/ and sound change in Australian English
- Ho-Hsien PAN
  - Lexical Tonal Space and Sandhi Rule
- Daniel PAPE
  - Intrinsic Pitch is not a universal phenomenon: Evidence from Romance languages
- Caterine PETRONE
  - Phonetic variation and phonological phrasing: does the Accentual Phrase exist in Italian?
- Nicolai PHARAO
  - Frequency of use and the Apparent Time Paradigm
- Janet PIERREHUMBERT
  - Lexical Bias in Cross-Dialect Word Recognition in Noise
- David PISONI
  - The role of phonetic detail in associating phonological units

- Kathy RASTLE
  - Can auditory distractors disrupt speech execution?
- Henning REETZ
  - An investigation of unaspirated dental and retroflex stops in Bangla
  - Harmful Reduction?
- Daylen RIGGS
  - The Scope of Phrasal Lengthening in Articulation: Prosody and Prominence
- Anastasia RIEHL
  - The Internal Structure of Nasal-Stop Sequences: Evidence from Austronesian
- Kevin ROON
  - Obligatory release and stiffness modulation in Moroccan Arabic
- Leslie ROOT
  - Distribution of tongue tip articulations in Hindi versus English and the acquisition of stop place categories

- Felix SCHAEFFLER
  - An investigation of cross-language differences in pitch range for speakers of English and German
- Jim SCOBBIE
  - Having your cake and eating it: An articulatory perspective on the individual’s place in systems of variation and change
- Michael SHELTON
  - The time course of stress and syllable-level encoding: Evidence from a delayed naming task in Spanish

- Chris SHEPPARD
  - The acquisition of lexical rhythm and duration by Japanese second language learners
- Ryan SHOSTED
  - The Acoustics of Voiceless Nasal Vowels
- E. Alleyn SMITH
  - Bringing Semantics to Sociophonetics: Social Variables and Secondary Entailments
- Caroline SMITH
  - Phrasing not duration differs in French questions and statements
- Rachel SMITH
  - Phonological learning based on interactive and mediated speech
- Paul SMOLENSKY
  - Replicator dynamics of speech perception and categorization
- Maria-Josep SOLE
  - Voicing control and nasalization
- Morgan SONDEREGGER
  - When does variation lead to change? A dynamical systems model of a stress shift in English
- Mee SONU
  - Vowel length effects on the perception of geminate and singleton stop boundaries by Korean learners of Japanese
- Shari SPEER
  - Processing lexical tone in third-tone sandhi
- Lorenzo SPREAFICO
  - Phonetic variation of /r/ in a language contact context: The case of South Tyrol Italian
- Ronald SPROUSE
  - Voicing control and nasalization
- Jane STUART-SMITH
  - Phonological learning based on interactive and mediated speech
  - Having your cake and eating it: An articulatory perspective on the individual's place in systems of variation and change
- Anita SZAKAY
  - Suprasegmentals and Ethnic Stereotypes in New Zealand

- Marija TABAIN
  - An EPG study of palatal consonants in Arrernte
- Terrin TAMATI
  - Lexical Bias in Cross-Dialect Word Recognition in Noise
- Phillip TIPTON
  - Here, [ðeː] and everywhere: lexical access, usage and the retrieval of non nonlinguistic
- Jane TSAY
  - Phonetic Details and Lexical Idiosyncrasies in Vowel Categorization in Taiwanese
- Hui-ying TSOU
  - Lexical Tonal Space and Sandhi Rule
- Michael TYLER
  - Greek-Australian bilinguals match the VOTs of Greek and Australian English native speakers depending on language context

- Akira UTSUGI
  - Ongoing merger of lexical tonal classes in Masan/Changwon Korean: Evidence from lexical and postlexical/intonational phenomena

- Alessandro VIETTI

- Phonetic variation of /r/ in a language contact context: The case of South Tyrol Italian
- Navin VISWANATHAN
  - Compensation for Coarticulation may reflect gestural perception: Evidence from a critical examination of the effects of non-speech contexts on speech categorization
- Travis WADE
  - Detailed Phonetic Memory for Multi-Word and Part-Word Sequences
- Abby WALKER
  - Phonetic information affects grammaticality judgements
  - Phonological, lexical, and frequency factors in coronal stop deletion in early New Zealand English
- Mary Ann WALTER
  - Vowel Cooccurrence in the Lexicon: An Anti-OCP Effect?
- Qian WANG
  - Do Chinese learners of English make use of fine phonetic details as English speakers do when perceiving the English lexical stress?
- Sarah WATSKY
  - The independence of auditory and categorical effects on speech perception
- Catherine WATSON
  - Ka conversion – the changing sound and rhythm of Māori?
- Laurence WHITE
  - Heads you lose: prosodic structure and timing in Hungarian
  - That elusive rhythm: Pros and cons of rhythm metrics
- Lukas WIGET
  - That elusive rhythm: Pros and cons of rhythm metrics
- Colin WILSON
  - Replicator dynamics of speech perception and categorization
- Wolfgang WOKUREK
  - Subglottal Resonances and Vowel Formant Variability: A Case Study of High German Monophthongs and Swabian Diphthongs
- Michael WOLMETZ
  - Replicator dynamics of speech perception and categorization
- Lei XU
  - Processing lexical tone in third-tone sandhi
- Tae-Jin YOON
  - Do Chinese learners of English make use of fine phonetic details as English speakers do when perceiving the English lexical stress?
- Ivan YUEN
  - Can auditory distractors disrupt speech execution?
- Chakir ZEROUAL
  - Obligatory release and stiffness modulation in Moroccan Arabic
- Frank ZIMMERER
  - Harmful Reduction?

Victoria
UNIVERSITY OF WELLINGTON
Te Whare Wānanga
o te Ūpoko o te Ika a Māui

Programme Abstracts Links, etc.

You are here: Home > Links

- **Home**
- **Abstracts**
  - View pdf versions of abstracts here
- **Programme**
  - Full programme
  - Poster sessions
- **Flyers & posters, Links, etc.**

Links, etc.

Downloadable posters advertising LabPhon11:

A3 size
A4 size

Previous LabPhon conferences:

LabPhon 10 - 2006 - Paris
LabPhon 9 - 2004 - University of Illinois at Urbana-Champaign
LabPhon 8 - 2002 - Yale University
LabPhon 7 - 2000 - University of Nijmegen
LabPhon 6 - 1998 - University of York
LabPhon 5 - 1996 - Northwestern University
LabPhon 4 - 1993 - Oxford University
LabPhon 3 - 1991 - University of California, Los Angeles
LabPhon 2 - 1989 - University of Edinburgh
LabPhon 1 - 1987 - The Ohio State University

LabPhon books

1. Kingston, John and Mary E. Beckman (1990, eds). *Between the grammar and physics of speech, Papers in laboratory phonology 1*. Cambridge: Cambridge University Press.
2. Docherty, Gerard J. and D. Robert Ladd (1992,eds). *Gesture, segment, prosody, Papers in laboratory phonology 2*. Cambridge: Cambridge University Press.
3. Keating, Patricia A. (1994, ed.). *Phonological structure and phonetic form, Papers in laboratory phonology 3*. Cambridge: Cambridge University Press.
4. Arvaniti, Amalia and Bruce Connell (1995, eds). *Phonology and phonetic evidence, Papers in Laboratory Phonology 4*. Cambridge: Cambridge University Press.
5. Broe, Michael and Janet Pierrehumbert (1999, eds). *Language acquisition and the lexicon, Papers in laboratory phonology 4*. Cambridge: Cambridge University Press.
6. Local, John, Richard Ogden, and Rosalind Temple (2003, eds). *Phonetic interpretation, Papers in Laboratory Phonology 6*. Cambridge: Cambridge University Press.
7. Gussenhoven, Carlos and Natasha Warner (2002, eds). *Laboratory Phonology 7*. Berlin: Mouton de Gruyter.
8. Goldstein Louis, D.H. Whalen and Catherine T. Best (2006, eds). *Laboratory Phonology 8*. Berlin: Mouton de Gruyter.
9. Cole, Jennifer and José Ignacio Hualde (2007, eds). *Laboratory Phonology*

9. Berlin: Mouton de Gruyter.

Victoria
UNIVERSITY OF WELLINGTON
Te Whare Wānanga o te Ūpoko o te Ika a Māui

Programme Abstracts Links, etc.

WELCOME TO THE

Laboratory Phonology 11 Conference

Home

Abstracts

View pdf versions of abstracts here

Programme

Full programme

Poster sessions

Flyers & posters, Links, etc.

where is Wellington? where is New Zealand? click on the map for more information

Additional sponsors:

UC UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

ASSTA
Australasian Speech Science & Technology Association

International Conference Fund
(Funded by the NZ Government and administered by the Royal Society of NZ)

Wellington, New Zealand
30 June - 2 July 2008

LabPhon 11 was held at Victoria University of Wellington, New Zealand, 30 June – 2 July 2008, and was followed at the same venue by the 5th International Gender and Language Association conference (IGALA5), 3-5 July 2008

The overall theme of the LabPhon 11 conference was ***Phonetic detail in the lexicon***, with the sub-themes shown below. Invited plenary speakers and commentators are listed against these sub-themes.

- **Accessing the lexicon**
  Invited speaker:**Keith Johnson, UC Berkeley**
  Commentator: **Marcus Taft, University of New South Wales**
  - How stable are lexical representations?
  - What encoding takes place between the signal and the lexicon?
  - To what degree does literacy/orthography influence lexical representation and access
- **Social information in the lexicon**:
  Invited speaker: **Paul Foulkes, University of York**
  Commentator: **Ben Munson, University of Minnesota**
  - Is phonetic information in the lexicon accompanied by social information?
  - How do social expectations about a speaker affect speech perception?
  - Is speaker-specific detail stored in the lexicon?
- **Phonetic cues to lexical structure**
  Invited speaker: **Alice Turk, University of Edinburgh**
  Commentator: **Laurie Bauer, Victoria University of Wellington**
  - To what degree does phonetic detail reflect morphological structure?
  - How are different types of boundaries expressed phonetically?
  - Is there phonetic evidence of lexical entries greater than the word?
- **Generalising over the lexicon**
  Invited speaker: **Anne Cutler, Max Planck Institute for Psycholinguistics & MARCS laboratory**
  Commentator: **Karen Croot, University of Sydney**
  - Is there evidence for or against the existence of abstract phonological categories?
  - What types of statistical generalisations emerge from the lexicon?
  - What is the nature of the relationship between the content of the lexicon and the shape of the phonological grammar?

Victoria University of Wellington
Te Whare Wānanga o te Ūpoko o te Ika a Māui

Poster sessions

Posters are listed below under each session, ordered alphabetically by first author surname. The poster display will be organised thematically, and poster presenters are asked to look for their name on a poster board in the display area. Posters for the Monday session can be put up any time on the Monday, and must be removed immediately after the poster session to make way for the posters for the Tuesday session, which can be put up at any time on the Tuesday, and should be removed after the poster session.

Monday 30th June 5.00-6.30	**Tuesday 1st July** 4.30-6.00
Adriaans, F. *Learning Phonotactic Generalizations from Continuous Speech: A Computational Study*	Albano, E. *Morphological and Phonetic Factors in Brazilian Portuguese Lexical Phonotactics*
Antoniou, M., Best, C.T., Tyler, M.D., & Kroos, C. *Greek-Australian bilinguals match the VOTs of Greek and Australian English native speakers depending on language context*	D'Imperio, M., & Petrone, C. *Phonetic variation and phonological phrasing: does the Accentual Phrase exist in Italian??*
Boll-Avetisyan, N., & Kager, R. *Identity Avoidance Between Non-Adjacent Consonants in Artificial Language Segmentation*	Docherty, G., Langstrof, C., & Foulkes, P. *Perceptual evaluation of sociophonetic variability: how do listeners learn?*
Brasileiro, I. *The Perception of a Vowel Contrast by Bilingual Children*	Felder, V., Joensson-Steiner, E., Eulitz, C., & Lahiri, A. *Experimental approaches to lexical representations of tone in Swedish: What is stored and how does it affect lexical word access?*
Brouwer, S., Mitterer, H., & Ernestus, M. *How acoustically reduced forms activate the lexicon: Evidence from eye-tracking*	Gibson, A. *Perception of Sung and Spoken Vowels in New Zealand English*
Brown, J., & Hansson, G. *Gradient phonotactics in the Gitksan lexicon*	Ingleby, M., & Ali, A. *Experiments on fine phonetic detail of Arabic syllable structure*
Calderone, B., Celata, C., & Herreros, I. *Recovering morphology from local phonotactic constraints*	Kaiser, E., & Munson, B. *Social Selectivity in Adults' Novel Sound Learning*
Dmitrieva, O., & Anttila, A. *The Gradient Phonotactics of English CVC Syllables*	Katseff, S., & Houde, J. *Compensation ?=? Mental Representation*
Finley, S., & Badecker, W. *Front/Back Asymmetries in Height Harmony*	Khattab, G., & Al-Tamimi, J. *Phonetic cues to gemination in Lebanese Arabic*
Guy, G.R., Hay, J., & Walker, A. *Phonological, lexical, and frequency factors in coronal stop deletion in early New Zealand English*	Kinoshita, N., & Sheppard, C. *The acquisition of lexical rhythm and duration by Japanese second language learners*
Hall, K.C. *Phonological Relationships: Beyond Contrast and Allophony*	Lahiri, A., & Reetz, H. *An Investigation of Unaspirated Dental and Retroflex Stops in Bangla*
Hu, F. *When the Ongoing Historical Sound Change Meets with Social Factors*	Li, F., Kong, E., Edwards, J., & Beckman, M. *Adult Acoustics and Developmental Patterns for Gender-marked Phonetic Variants in Mandarin Fricatives and Japanese Stops*
Kainada, E. *Influence of prosodic phrasing on stop voicing and vowel hiatus resolution in Modern Greek*	Madsack, A., Lulich, S.M., Wokurek, W., & Dogil, G. *Subglottal Resonances and Vowel Formant Variability: A Case Study of High German Monophthongs and Swabian Diphthongs*
Kapatsinski, V., & Pisoni, D. *The role of phonetic detail in associating phonological units*	Mok, P.K., & Hawkins, S. *Syllabifications of the /st/ cluster and vowel-to-vowel coarticulation in English*
Kingston, J., Kawahara, S., Chambless, D., Key, M., &	Morley, R.

Watsky, S. *The independence of auditory and categorical effects on speech perception*	*Integrating Sub-phonemic Cues: learning across a morpheme boundary*
Loakes, D. *Phonetic Evidence for Neutralisation of Prelateral /æ/ and /e/ in Melbourne English*	Mücke, D., Grice, M., Hermes, A., Becker, J., & Baumann, S. *Articulatory Expression of Focus in German*
Maddieson, I. *Segmentation, not just similarity: alliteration as a natural experiment*	Nagy, N. *Production Frequency and Formant Frequency in R speech*
Mennen, I., Schaeffler, F., & Docherty, G. *An investigation of cross-language differences in pitch range for speakers of English and German*	Niyogi, P., & Sonderegger, M. *When does variation lead to change? A dynamical systems model of a stress shift in English*
Pan, H., Li, P., Lee, C., Huang, H., & Tsou, H. *Lexical Tonal Space and Sandhi Rule*	Ogasawara, N. *Does what listeners hear affect what they say? Evidence from shadowing experiment*
Pharao, N. *Frequency of use and the Apparent Time Paradigm*	Pape, D., & Mooshammer, C. *Intrinsic pitch is not a universal: Evidence from Romance languages*
Roon, K., Gafos, A.I., Hoole, P., & Zeroual, C. *Obligatory Release and Stiffness Modulation in Moroccan Arabic*	Scobbie, J., Stuart-Smith, J., & Lawson, E. *Having your cake and eating it: An articulatory perspective on the individual's place in systems of variation and change*
Shelton, M., Gerfen, C., & Gutiérrez Palma, N. *The time course of stress and syllable-level encoding: Evidence from a delayed naming task in Spanish*	Smith, C. *Phrasing not Duration Differs in French questions and statements*
Shosted, R. *The Acoustics of voiceless nasal vowels*	Tipton, P. *Here, there and everywhere: lexical access, usage and the retrieval of non-linguistic information in language production*
Solé, M-J., Sprouse, R., & Ohala, J. *Voicing control and nasalization*	Tsay, J., & Lee, Y. *Phonetic Details and Lexical Idiosyncrasies in Vowel Categorization in Taiwanese*
Sonu, M. *Vowel length vowel effects on the perception of geminate and singleton stop boundaries by Korean learners of Japanese*	Utsugi, A. *Ongoing merger of lexical tonal classes in Masan/Changwon Korean: Evidence from lexical and postlexical/intonational phenomena*
Szakay, A. *Suprasegmentals and Ethnic Stereotypes in New Zealand*	Vietti, A., & Spreafico, L. *Phonetic variation of /r/ in a language contact context: The case of South Tyrol Italian*
Tabain, M. *An EPG study of palatal consonants in Arrernte*	Viswanathan, N., Magnuson, J.S., & Fowler, C.A. *Compensation for Coarticulation may reflect gestural perception: Evidence from a critical examination of the effects of non-speech contexts on speech categorization*
Watson, C., Harlow, R., Bauer, W., Maclagan, M., King, J., & Keegan, P. *Ka conversion – the changing sound and rhythm of Māori?*	Walker, A. *Phonetic information affects grammaticality judgments*
White, L., Mattys, S., & Grenon, I. *That elusive rhythm: Pros and cons of rhythm metrics*	Wang, Q., & Yoon, T.J. *Do Chinese learners of English make use of fine phonetic details as English speakers do when perceiving the English lexical stress?*
Yuen, I., Rastle, K., & Brysbaert, M. *Can auditory distractors disrupt speech execution?*	White, L., & Mády, K. *Heads you lose: prosodic structure and timing in Hungarian*
Zimmerer, F., Lahiri, A., & Reetz, H. *Harmful Reduction?*	Wilson, C., Wolmetz, M., & Smolensky, P. *Replicator dynamics of speech perception and categorization*

Victoria
UNIVERSITY OF WELLINGTON
Te Whare Wānanga
o te Ūpoko o te Ika a Māui

Search

Search results are currently displayed on the main Victoria website. If you wish to return to this site, please use your browser back button.

Search

This Site   All Victoria Sites

>

Other Options:

- A-Z List of all Victoria Sites
- Victoria Glossary

Learning Phonotactic Generalizations from Continuous Speech: A Computational Study

Frans Adriaans

Utrecht Institute of Linguistics OTS, Utrecht University, the Netherlands; frans.adriaans@let.uu.nl

Infants acquire knowledge of phonotactics well before the end of the first year of life (e.g., Jusczyk, Friederici, Wessels, Svenkerud, & Jusczyk, 1993) and use phonotactics in the segmentation of continuous speech (Mattys & Jusczyk, 2001). This suggests that models of phonotactic learning should consider continuous speech as a source of data through which phonotactics can be acquired. Existing models of phonotactic learning, however, assume that phonotactic constraints are induced from the lexicon (Pierrehumbert, 2003; Hayes & Wilson, to appear). What learning mechanisms could be involved in *pre*-lexical phonotactic learning?

Saffran, Aslin, & Newport (1996) show that infants are sensitive to transitional probabilities of adjacent syllables when learning artificial words from a stream of continuous speech. Since a low transitional probability indicates a likely word boundary, statistical learning helps in creating a lexicon from continuous speech input. Learners are also sensitive to statistical dependencies between segments (Newport & Aslin, 2004). These studies suggest a possible role for statistical learning in the acquisition of pre-lexical phonotactics. There is evidence, however, that infant phonotactic acquisition is not limited to the co-occurrences of specific segments, but involves the learning of restrictions on the co-occurrences of natural classes. Saffran & Thiessen (2003) show that 9-month-old infants are able to learn patterns of voicing, but not patterns of unrelated segments. The infant needs to take into account the similarity between segments in order to learn such phonotactic generalizations.

The present paper addresses two issues, taking a computational angle. First, we investigate whether it is possible to induce phonotactic constraints from transcriptions of continuous speech. Second, we empirically determine the role of phonotactic generalizations in speech segmentation. More specifically, we predict that phonotactic constraints, which are induced through a combination of segment-based statistical learning and feature-based generalization, provide a more reliable cue for detecting potential word boundaries than segment-based statistical learning without generalization. Phonotactic generalizations, learned from continuous speech, may thus enhance the learner's ability to segment speech.

We propose a computational model, which links statistical learning to generalization. As input to our model, we use the broad phonetic transcriptions of the Spoken Dutch Corpus (Goddijn & Binnenpoorte, 2003), where all word boundaries within utterances have been removed. The learner obtains biphone observed/expected (O/E) ratios through statistical learning. The learner then categorizes biphones into two types of constraints (each defined by their own O/E thresholds), which serve as a basis for generalization:

(1) **xy* ‘Sequence *xy* should not occur word-internally.'
CONTIG-IO(*xy*) ‘Sequence *xy* should occur word-internally.'

The learner creates more general constraints from the statistically learned biphone constraints using a generalization algorithm. This algorithm takes advantage of the similarity between the biphone constraints, which is quantified as the number of shared features[1]. We follow the definition of constraint neighbors, proposed in Hayes (1999): Two constraints are neighbors if they have different values for one single feature. Our generalization algorithm creates a new, more general, constraint where this feature has been removed. The resulting constraints cover sequences of natural classes, instead of sequences of segments.

As a result of generalization, the learner has to deal with overlapping, and possibly conflicting constraints in speech segmentation. For example, the general constraint CONTIG-IO([*plosive*][*liquid*]) states that no boundary should be inserted into a sequence of a plosive followed by a liquid. This conflicts with the biphone constraint *tl, which says that ‘tl' should not occur word-internally. We borrow the principle of

[1] Our implementation uses the features *syl*, *cons*, *appr*, *son*, *cont*, *nas*, *voice*, *place*, *ant*, *lat* for consonants, and *high*, *low*, *back*, *round*, *long*, *tense*, *nasd* for vowels.

strict domination from Optimality Theory (Prince & Smolensky, 1993) in order to resolve conflicts between the learned phonotactic constraints. The constraint ranking is based on the values obtained through statistical learning. The ranking captures segment-specific exceptions to feature-based generalizations.

We conducted a series of computer simulations, which aim at comparing purely statistical models, such as transitional probability, to our constraint induction model, which combines statistical learning and generalization. The models are tested on their ability to detect word boundaries in the Spoken Dutch Corpus, using test sets that were withheld from training. The task of the models is to maximize the number of hits (correct boundaries), while minimizing the number of false alarms (incorrect boundaries). The d-prime score expresses how well the model performs in distinguishing hits from false alarms.

Model	Hit rate	False alarm rate	d-prime
1. Random	0.5012	0.4991	0.0051
2. Statistical learning (SL)	**0.6108**	0.2241	1.0399
3. Statistical learning (SL) + generalization	0.3681	**0.0747**	**1.1060**

Table 1: Simulation results for three different models. Model 1 uses random segmentation, which involves no learning. Model 2 uses segment-based statistical learning (transitional probability). Model 3 is the constraint induction model, which, in addition to segment-based statistical learning, uses feature-based generalization.

The results in Table 1 show that our model outperforms statistical learning (as reflected in the d-prime scores). In addition, our model is conservative: It places fewer, but more reliable boundaries than the SL model. Such a strategy results in accurate proto-words (e.g. *Thisis-thedoggy*), whereas oversegmentation strategies (in this case, the SL model) result in inaccurate lexical entries (e.g. *This-is-the-do-ggy*).

To conclude, our model learns phonotactic constraints from continuous speech through generalization over statistically learned biphone constraints. The simulations show that the combined strengths of statistical learning and generalization provide the learner with a more reliable cue for detecting word boundaries in continuous speech than statistical learning alone (i.e. without generalization).

References

Goddijn, S. & Binnenpoorte, D. (2003). Assessing manually corrected broad phonetic transcriptions in the Spoken Dutch Corpus. In M. J. Solé, D. Recasens & J. Romero (Eds.), *Proceedings of 15th International Congress of Phonetic Sciences* (pp. 1361-1364). Rundle Mall: Causal Productions.

Hayes, B. (1999). Phonetically driven phonology: The role of Optimality Theory and inductive grounding. In M. Darnell, E. Moravscik, M. Noonan, F.J. Newmeyer & K.M. Wheatley (Eds.), *Functionalism and Formalism in Linguistics* (pp.243–285). Amsterdam: John Benjamin.

Hayes, B. & Wilson, C. (2007). A maximum entropy model of phonotactics and phonotactic learning. (to appear).

Jusczyk, P. W., Friederici, A. D., Wessels, J. M. I., Svenkerud, V. Y., & Jusczyk, A. M. (1993). Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language*, 32, 402–420.

Mattys, S. L. & Jusczyk, P. W. (2001). Phonotactic cues for segmentation of fluent speech by infants. *Cognition*, 78, 91–121.

Newport, E. L. & Aslin, R. N. (2004). Learning at a distance: I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, 48, 127–162.

Pierrehumbert, J. B. (2003). Probabilistic phonology: Discrimination and robustness. In R. Bod, J. Hay & S. Jannedy (Eds.). *Probabilistic Linguistics*. Cambridge, MA: MIT Press.

Prince, A. & Smolensky, P. (1993). *Optimality theory: Constraint interaction in generative grammar*. Technical report, Rutgers University Center for Cognitive Science.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274(5294), 1926–1928.

Saffran, J. R. & Thiessen, E. D. (2003). Pattern induction by infant language learners. *Developmental Psychology*, 39(3), 484–494.

Morphological And Phonetic Factors In Brazilian Portuguese Lexical Phonotactics

Eleonora C. Albano

Laboratório de Fonética e Psicolingüística, State University of Campinas, Brazil;
albano@unicamp.br

Overview

This paper approaches Brazilian Portuguese (henceforth BP) phonotactics by examining co-occurrence frequencies of stressed consonant-vowel pairs in oral and written vocabulary. The aim is to throw light on the interplay between phonetic and morphological factors in shaping probabilistic phonotactics.

The materials were compiled from phoneticized speech transcripts and newspaper texts. Three data sets are compared for each lexical database: the total 'CV frequencies from the full word list; a subset thereof from which words with frequent and/or productive suffixes have been removed; and another subset thereof restricted to word initial position. The statistical technique employed is correspondence analysis (Benzécri 1980).

The results point to a striking coherence between the oral and the written vocabulary. Both show a tendency for the plots to split into phonetically interpretable regions, but, in the full lists, morphological effects perturb an otherwise straightforward interaction between segment frequency and phonetic content.

Method

Two publically available corpora were used to extract the word types: a newspaper corpus (about 35,000 relevant words), and an orthographically transcribed oral language corpus (about 30,000 relevant words). The lists were derived from running text without lemmatization. Orthography to phone conversion was achieved by a program which treats ambiguous graphemes with 98% success (Albano & Moreira 1996). The resulting transcriptions are broad, but differentiate positional allophones. 'CV co-occurrence frequencies were thus computed for each version of each lexicon.

Like European Portuguese, BP has only one stressed syllable per word, except for compounds. Stress may fall on the antepenult, the penult – which is the majority, default case (about 55%) –, or the final (about 35%). Stress coincides with the stem end, except where antepenultimate (about 10%). Derivational suffixes automatically shift stress to the right.

Only simple onsets are included in the analysis. The 133 target 'CV pairs result from the combination of the 19 medial consonants, /p, b, m, f, v, t, d, n, s, z, l, ɾ, ɲ, ʎ, ʃ, ʒ, k, g, ʀ/, with the 7 stressed vowels, /i, e, ɛ, a, ɔ, o, u/. The initial consonant inventory is reduced to 16 (/ɾ, ɲ, ʎ/ are absent). Stressed syllables with complex rhymes are included, since the coda is not at stake. For example, so-called nasal vowels and diphthongs are tallied with their oral counterparts in accordance with the transcription, which represents nasalization with a following underspecified nasal.

The most frequent derivational suffixes in both the oral and the written database are[1]: *'a/ição, ões'*; *'a/idor, a'*; *'amente'*; *'a/imento'*; *'aria'*; *'a/isão, ões'*; *'icar'*; *'idade'*; *'ificar'*; *'(i)tor, a'; 'izar'*; *'ogia'; 'ogismo'*; *'onal'*; *'onar'*; *'zinho, a'*. These suffixes account for 10 to 15% of the databases. 'CV biases due to suffixation are thus filtered by ignoring words containing them.

Correspondence analysis is a non-parametric multivariate exploratory technique which extracts orthogonal factors from chi-square values. In dealing with lexical frequencies, it is often preferable to principal component analysis, which requires transformation of the raw data to fit the normal distribution – a hard, if not impossible, requirement to meet with nominal data. Effect sizes are evaluated through Cramer's V. Though not strong, effects are far from negligible ($V \geq .10$), and stable across lexicons. Since most inferences permitted by correspondence analysis are qualitative – i.e., rely on groupings in the coordinate

[1] Some suffix initial vowels vary according to inflection class.

space, effect sizes, etc. –, O/E ratios for all six contingency tables were also calculated, in order to explore and rank specific biases.

Results

The Cartesian plots extracted from the 'CV frequencies in the full vocabulary, oral or written, show some strong effects of vowels and consonants occurring in certain derivational suffixes. Such segments drift away from the groupings in their quadrants. The most favored combinations are: (1) /'me/, due to *'a/imento'*, which forms nouns from verbs, in addition to *'amente'*, which forms manner adverbs from adjectives; and (2) /'do/, due to *'a/idor'*, which forms agentive nouns from verbs. A lesser, but still appreciable bias, favors /'sa/, and is due to the frequency/productivity of *'a/ição'*, which forms abstract nouns from verbs. The first two coordinates, where phonetic and morphological effects are confounded, account for 60 to 70% of the inertia. The third coordinate accounts for another 10%, and brings out some further phonetics/morphology interactions.

In turn, the Cartesian plots derived from the 'CV frequencies in the morphologically filtered vocabulary, oral or written, tend to split into regions that lend themselves to easy phonetic interpretation, corresponding, roughly, to: (1) round, front, and back, for vowels; and (2) labial, coronal and velar, for consonants. When coordinates are scaled so that vowels and consonants are plotted together, the first two coordinates, which account for 70 to 80 % of the inertia, can be interpreted as arising from the interaction between segment frequency and constriction location – which is, by far, the clearest and strongest grouping bias. There are grounds for the inference – further supported by the O/E ratios – that C's and V's tend to agree in constriction location. Additional coordinates are not worth extracting in view of their small loadings and opaque patterning.

The Cartesian plots derived from the 'CV frequencies in initially stressed words reproduce the trends observed in the morphologically filtered vocabulary, corroborating the hypothesis that derivational suffixes are responsible for masking the bias for constriction location agreement in both vocabularies.

O/E ratios greater than 1.5 give a further glimpse into the underlying logic of BP phonotactics. The highest ranks involve the above mentioned suffixes *'a/imento'*, *'amente'*, and *'a/idor'*. Note that both /'me/ and /'do/ violate constriction location agreement, i.e., the 'CV pair is, rather, contrastive. The next highest ranks involve an intriguing asymmetry among labial obstruents: /b, p, f/ prefer /ɔ, u/, while /v/ prefers /e, ɛ/. The remaining high ranks are related to a negative bias: velars show a slight preference for back vowels along with a strong rejection for front vowels (O/E ≤ .5).

Conclusion

These results suggest that both speech production – as reflected in constriction location agreement – and speech perception – as reflected in constriction location contrast – play a role in structuring BP lexical phonotactics.

Acknowledgements:

Support is gratefully acknowledged from *CNPq,* the Brazilian federal research agency (grant no. 311024/2006-8), and *FAPESP,* the São Paulo State research agency (grant no. 01/00136-2).

References:

Albano, E. C. & Moreira, A. A. (1996). Archisegment-based letter-to-phone conversion for concatenative synthesis in Portuguese. *Proceedings ICSLP'96*, 3, 1708-1711.

Benzécri, J.-P. (1980[1973]). *L'analyse des données, tome 2: l'analyse des correspondances*. Paris: Bordas.

Gradience in Morphological Binding: Evidence from Perception of Audiovisually Incongruent Speech

Azra N. Ali & Michael Ingleby

School of Computing and Engineering, University of Huddersfield, England;
a.n.ali@hud.ac.uk m.ingleby@hud.ac.uk

Over the last three decades, empirical studies have established that the perception of a target word, whether presented visually as text or spoken as an acoustic signal, is influenced by the morphology of the target. Generally, targets have been subject to masking and/or priming. Primes for perception of a spoken word, for e.g. have been lexical or pictorial with some morphological similarity to the target. These types of experiment are designed supposing that perception of the prime and the separate perception of a target are mediated by the same structural model of lexical items, even when the prime and target engage different channels of perception. To escape this supposition we have sought for an empirical probe that engages only one kind of processing – that for natural audiovisual speech signals. We use dubbed speech, but guarantee ecological validity by placing dubbed stimuli at random amongst natural stimuli. In such natural context, incongruent stimuli, in which the audio and visual signals are in conflict, elicit from many a perception that differs from the data in either channel. The phenomenon is known as McGurk fusion (McGurk & MacDonald, 1976). Typically audio 'map' temporally aligned with visual 'mack' (lip movements) elicits the fusion percept 'mat', or more symbolically

(1) $_{\mathrm{AUD}}(\textit{ma}\boldsymbol{p} \parallel \textit{ma}\boldsymbol{ck})_{\mathrm{VIS}} \rightarrow (\textit{ma}\boldsymbol{t})_{\mathrm{PER}}$

and (2) $_{\mathrm{AUD}}(\boldsymbol{sp}\textit{ill} \parallel \boldsymbol{sk}\textit{ill})_{\mathrm{VIS}} \rightarrow (\boldsymbol{st}\textit{ill})_{\mathrm{PER}}$

and also (3) $_{\mathrm{AUD}}(\textit{ma}\boldsymbol{ps} \parallel \textit{ma}\boldsymbol{cks})_{\mathrm{VIS}} \rightarrow (\textit{ma}\boldsymbol{ts})_{\mathrm{PER}}$

The proportion of the participants that report fusion ('the fusion rate'), depends on the morphological environment. In example (1), the site is the syllabic coda of a stem; while the site is in the second branch of a branching syllabic onset in (2) and the first branch of a branching coda in example (3) above. An abiding pattern for fusion rates is that they are significantly higher at coda sites than at onset sites, and this has been used to probe the syllabic structure of English and Arabic words (Ali *et al*, 2005).

The purpose of this new study is to investigate the impact and gradience of four of the most frequent affixes, such as *-ing, -s, -er,* and *–y/ie* on fusion rates at a stem site. In (3), for example, a stem coda site is sited before an inflectional *-s*. Generally the effect of affixing lowers the fusion rate: measured rates for examples (1) and (3) are respectively ~80% and ~50%, a lowering of about 30 percentage points with affixing in (3). A small corpus of incongruent audiovisual stimuli was tested in fusion experiments. The results showed that the fusion rates at stem coda sites were not lowered by the same amount by different morphological suffixes, as shown in the schematic distribution of the fusion data presented in Figure 1. The distribution curves are made by fitting rates for a number of different stems to a Gaussian distribution, and show that there appears to be gradience amongst the four exemplar morphological suffixes: *–ing* has mean (m) 69%, fraction lowering (FL) 10.08%, *–y* has m=63%, FL=16.84%, *–s* has m=54%, FL=30.09% and *–er* has m=43%, FL=45.02%. It appears that certain morphological affixes bind to the foregoing stem and lower the bare-stem fusion rates (80%) more weakly than others. Statistical tests confirm that there are significant differences in the binding effects of the different affixes – thus gradience in the morphological binding. If we nominate a decision threshold of, say, 25% for the fractional lowering of fusion rate by a suffix (Figure 1), then there appear to be two categories of binding: weak binding in Group 1 {*-ing* and *–y*} and strong binding in Group 2 {*-s* and *–er*}. Further, statistical analysis (t-test) reveals significant differences between the two groups but no significant differences within each group. This statistically-based categorical pattern is not the grammarian's distinction between inflectional and derivational binding: inflectional *–ing* and derivational *-y* bind weakly, while derivational *–er* and inflectional *–s* bind strongly. There is growing evidence from other linguistic phenomena that there are two levels of binding strength: strong or analytic binding, in which stem and particle form an indivisible morphological simplex, *versus* weak or non-analytic binding, which allows stem and affix some degree of separation in a complex. For example, the English stems *húmid* and *párent* keep their stress patterns under weak derivational binding to form *húmidly* and *párenthood* but change them under strong but still derivational binding to form *humídity* and *paréntal*. Similar stress phenomena are found in very different languages such as Arabic. Furthermore, in Germanic languages, the final consonant of a morphological domain is devoiced: the stem word *Kind* ('child') devoices to [kɪnt] even when bound to adjectival particle *–lich,* to form *Kindlich,* (pronounced [kɪntlɪx]) but, when bound strongly by genitive inflection *–es* or adjectival derivation *-isch* , the domains associated with *Kindes* and *Kindisch* are simplex and their non-final *d* is voiced (Brockhaus, 1995).

Studies of perception errors (Janssen & Humphreys, 2002) indicate that inflectional suffixes are more susceptible to perceptual errors than derivational suffixes; there were more perceptual errors when *–er* was a real morpheme than when it was a pseudo-morpheme as in 'summer', 'corner', etc.. Our stimuli consisted of only real morpheme *–er* words, and we found bounding gradience did not respect grammatical categories. It is possible that binding strength in morphology is an habituation phenomenon. There are indications that relative frequency induces gradience morphological decomposability (Hay, 2003). For example, if a derived word is more frequent than its stem word, then the derived word is likely to be accessed as a whole. If the derived word is less frequent than its base then it is likely to be more complex and thus decomposed. In our fusion experiments, the stimuli consisted of words with high stem frequencies and low word frequencies in the British National Corpus. Thus, our stimuli fall into Hay's salience regime for high decomposability. Nevertheless, within this regime our results show a gradience of binding: even with the simple exemplar suffixes of the study, our participants tended to process some words as complex units and others as simplices.

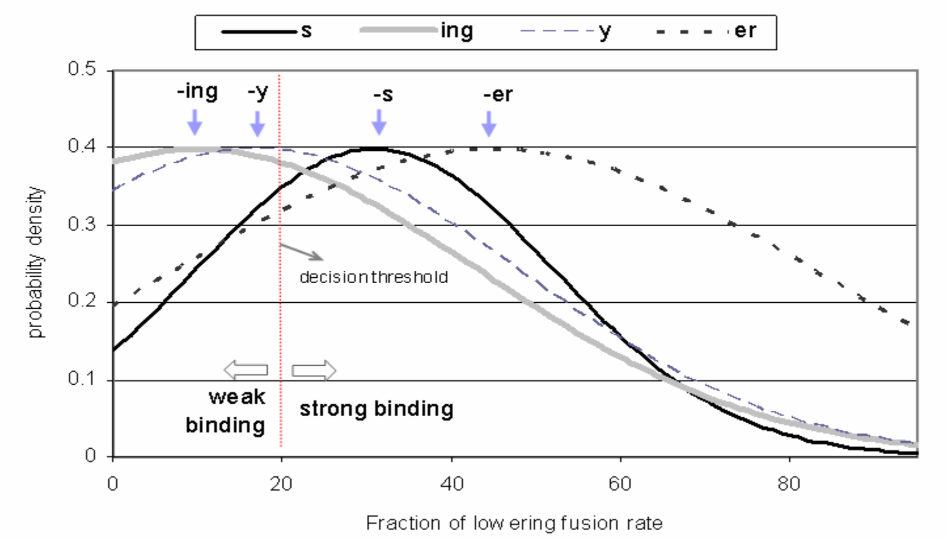


Figure 1: Schematic distribution of fraction of lowering of fusion rates.

References

Ali, A.N., Hassan-Haj, A., Ingleby, M. & Idrissi, A. (2005). McGurk fusion effects in Arabic words. *International Conference of Audio Visual Speech Processing*, Vancouver, Canada, 11-16.

Brockhaus, W. (1995). *Final Devoicing in the Phonology of German*. Tübingen: Max Niemayer Verlag.

Janssen, D. & Humphreys, K. (2002). Morphological speech errors on agentive and comparative affixes. Poster presented at the *Third International Conference on the Mental Lexicon*. Banff, Canada.

Hay, J. (2003). *Causes and Consequences of Word Structures*. London: Routledge.

McGurk, H. & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264, 746-748.

Plasticity in voice category learning

Attila Andics[1,2] & James M. McQueen[1]

[1]Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands
[2]FC Donders Centre for Cognitive Neuroimaging, Radboud University Nijmegen, Netherlands
attila.andics@mpi.nl, james.mcqueen@mpi.nl

Behavioural and neuroscientific studies indicate that voice and speech processing are partly independent, but interact at an early stage (e.g., Knösche, Lattner, Maess, Schauer, & Friederici, 2002). As we learn to identify new talkers by listening to the words they speak, we are exposed to voice-specific non-segmental and segmental information. Which type contributes to voice category formation, and how plastic are these categories? We examined this question in a two-day voice category training experiment. We asked whether listeners are able to generalize knowledge about a voice to untrained tokens of trained words and to words with no segmental overlap, and whether they are able to re-learn boundaries of individual voice categories.

Sixteen Dutch listeners were presented with morphed continua between two unfamiliar male voices. They made forced-choice decisions on speaker identity (voice A or voice B). Intermediate steps were made by morphing natural endpoint tokens, using STRAIGHT (Kawahara, 2006). There were two 18-minute training phases on both days, each preceded and followed by a 9-minute long test. The pretest on Day 1 served as a baseline; that on Day 2 provided a measure of consolidation. During training, a continuum with the word [mɛs] was used. The category boundary was made explicit by giving feedback according to a predefined boundary at 50% voice B morphs one day (symmetric training) and at either 30% or 70% the other day (asymmetric training). This training was amplified by presenting more stimuli from the most ambiguous part of the continuum: the mean of all stimuli from a voice category was a 10% distance from the category boundary. The whole continuum was sampled both during training and at test, but there was no stimulus overlap between the two parts. At test no feedback was given. Three continua were tested: the trained [mɛs] continuum, another [mɛs] continuum based on different tokens from the same speakers, and a [lot] continuum from the same speakers. Training phases contained 360 trials (12 repetitions of 30 morph levels). Test phases contained 270 trials (10 repetitions of 9 morph levels on 3 continua).

The plasticity of listeners' perceptual systems in response to different training conditions was first investigated in within-subject paired t-tests. Symmetric (50%) versus asymmetric (30% or 70%) boundary training affected categorization differently: the identification curve shifted in accordance with the trained boundary shift (Fig.1a-c). This suggests that listeners were able to learn and re-learn voice category boundaries with a one day delay. A repeated-measures ANOVA examined the effect of voice familiarity (unfamiliar as measured at baseline, minimal familiarization, i.e. 18 mins, and maximal familiarization, i.e. 72 mins over two days). We found a significant interaction of familiarity with morph level: Familiarization made the categorization curves steeper (Fig.1d). We then contrasted the asymmetric training conditions between subjects. Response times were shortest for the voice endpoints and longest for the ambiguous stimuli. We found a significant training by level interaction: The difference in the position of the RT peak across groups suggests a difference in the perceived category boundary (Fig.1e). We also tested training and word effects (trained [mɛs], untrained [mɛs], untrained [lot]) on categorization responses. We found significant main effects of training and word, and significant interactions: training by word, training by level, training by session, and word by level. The training effect was largest for the word [mɛs], for the morph levels around the boundary, and for the test session immediately after 36 mins training. A sessionwise, word-by-word analysis showed that the training-related boundary shift effect was present (1) already after 18 mins training for the untrained [mɛs] continuum, (2) but only after 36 mins training for all tested word continua, including the segmentally not overlapping word [lot], and (3) even 24 hours after training for both the trained and untrained [mɛs] continua. These results suggest that non-segmental voice-specific information was used for voice identification. The main effect of word was most expressed around levels corresponding to the category boundaries for the three word continua, suggesting that categorization also involved segmental information. The fact that the boundary shift was larger for untrained [mɛs] than for [lot] also shows the use of segmental information (Fig.1b,c).

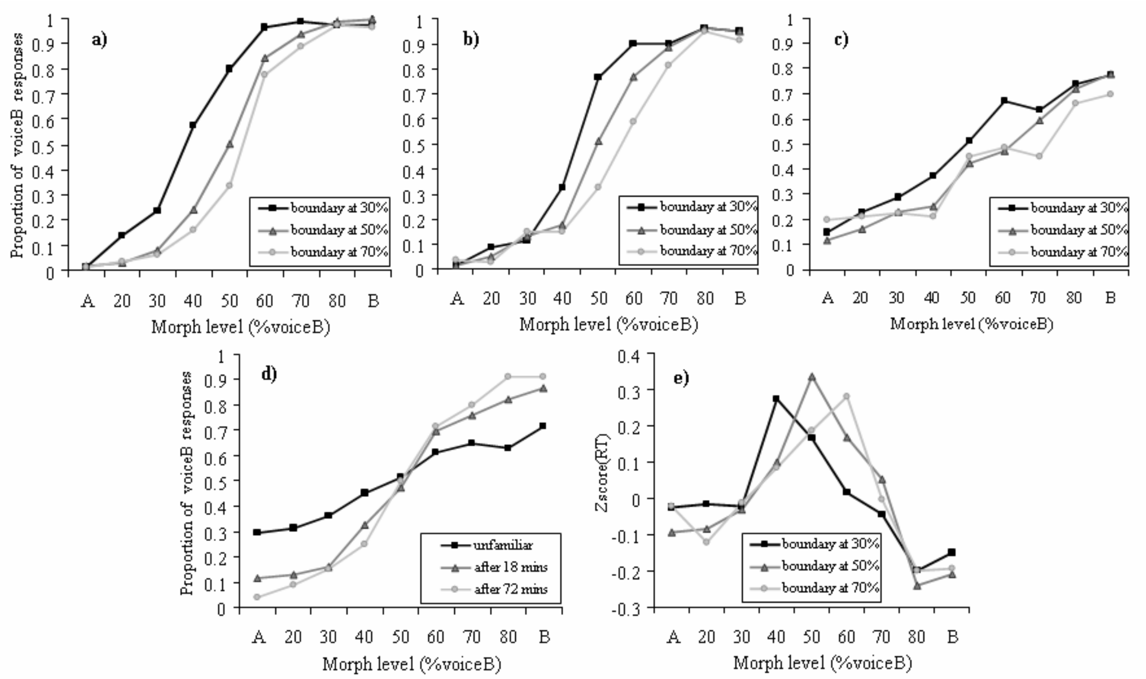


Figure 1: Voice categorization performance. a-c) Voice categorization responses per training group after 36 mins training for a) the trained [mɛs], b) the untrained [mɛs] and c) the untrained [lot]. d) Voice categorization responses collapsed across training groups, after different amounts of training. e) RTs per training group after 36 mins training.

Perception of the voice morph continuum thus became more categorical with training, and, as a function of the changes in the boundary settings, identification curves were shifted. Listeners appear able to set and reset voice category boundaries during training with feedback. These changes take place quickly and last long. This category boundary shift generalizes to non-trained utterances (1) of the same continuum, (2) of a different continuum with the same word, and (3) even to a different word with no segmental overlap. This shows that the training leads to general knowledge about the voices that is not specific to individual segments. Voice identity judgments differed for different word continua, however, suggesting that voice knowledge also includes segmental information. We conclude that voice categories are plastic and abstract, and are based at least in part on local cues in the speech signal.

References

Kawahara, H. (2006). STRAIGHT, exploitation of the other aspect of VOCODER: Perceptually isomorphic decomposition of speech sounds. *Acoustical Science and Technology*, 27(6), 349-353.

Knösche, T.R., Lattner, S., Maess, B., Schauer, M., & Friederici, A.D. (2002). Early parallel processing of auditory word and voice information. *NeuroImage*, 17, 1493-1503.

Greek-Australian bilinguals match the VOTs of Greek and Australian English native speakers depending on language context

Mark Antoniou[*], Catherine T. Best[*#], Michael D. Tyler[*] & Christian Kroos[*]

[*]MARCS Auditory Laboratories, University of Western Sydney, Australia;
[#]Haskins Laboratories, New Haven, CT, USA
m.antoniou@uws.edu.au, c.best@uws.edu.au, m.tyler@uws.edu.au, c.kroos@uws.edu.au

Fluent bilinguals frequently adjust their speech to match the linguistic setting. Linguistic context effects impact on the bilingual's selection of language-specific lexical items, morphological units, and syntactic settings when speaking. We would expect such contextual effects may appear even in the phonetic settings of bilinguals' speech production where those differ between their languages. While context effects have been addressed in theories of bilingual word selection, lexical and syntactic code switching and other higher-order aspects of language use (Green, 1998; Grosjean, 2001), they have barely been touched by theories of phonetics and phonology. For example, the Speech Learning Model (Flege, 1995), which attempts to predict foreign-accented speech in second language learners, says nothing about whether/how bilinguals will shift their production of speech depending on the linguistic context. Moreover, only a few cross-language studies have investigated the influence that linguistic context can exert on bilingual speech production (Caramazza, Yeni-Komshian, Zurif, & Carbone, 1973; Flege & Eefting, 1987; Hazan & Boulakia, 1993; Magloire & Green, 1999). Those few have provided mixed results, possibly due to methodological differences and limitations.

The present study addressed the effects of linguistic context in bilingual speech production in a carefully controlled study on stop consonants in Australian English (AusE) and Modern Greek (MG), spoken by numerous bilinguals in Sydney. These languages make stop voicing distinctions at different points along the voice onset time (VOT) dimension. For example, in word-initial position, voiceless stops /p, t/ have short lag unaspirated VOT in MG ([p, t]) but long lag aspirated VOT in AusE ([p^h, t^h]), and voiced stops /b, d/ are prevoiced in MG ([b, d]) but typically have short lag unaspirated VOT in AusE ([p, t]).

To address whether and to what degree these bilinguals modify their VOT values according to linguistic context, we systematically controlled the language of presentation with the participants in two subgroups: a Greek mode subgroup (GM) and an English mode subgroup (EM). The language of presentation was manipulated between groups to optimise the likelihood that the bilinguals would adjust their speech in a language-specific way according to the linguistic setting. It was hypothesised that they would produce different VOTs depending on the language of presentation, equivalent to those of monolingual speakers in the corresponding linguistic context. Specifically, it was predicted that GM bilinguals would match the VOTs of MG monolinguals, and EM bilinguals would match the VOTs of AusE monolinguals.

Two groups of fluent Greek-Australian English early sequential bilinguals (Sydney, Australia) were compared with groups of AusE (Sydney) and MG native speakers (Athens, Greece). All bilingual participants had been exposed to MG since birth, acquiring it as their first language, and later acquired English in Australia before the age of six years. They continue to communicate in MG in their everyday lives, but have become dominant in AusE, their second language.

Speakers produced the voiced stop consonants /b, d/ and voiceless stops /p, t/ in mono- and bi-syllabic (trochaic and iambic) nonwords (/Ca/, /'aCa/, /a'Ca/) in carrier sentences that were presented visually on a computer monitor in random order. Four tokens of each target were recorded. All contact, instructions, forms, carrier phrases, and feedback occurred in only one language for a given participant (MG or AusE). The MG and AusE monolinguals received all information and forms in their native language.

The recordings were segmented and the target word stop VOTs were acoustically analysed using Praat. Stops were classified as prevoiced (negative VOT) only if voicing was present just prior and during release. If voicing was absent immediately preceding release, the stop was classified as voiceless (positive VOT).

Separate ANOVAs were conducted for each stop consonant (/b, p, d, t/), in each positional context. In initial position, there was a significant language of presentation difference in VOT for the voiced bilabial /b/

$F(3, 25) = 37.777$, $p < .001$, as it was produced with voicing lead by the MGs (Mean = -124ms) and GM bilinguals (-132ms) but with nearly 0 VOT by the AusEs (0.0001ms) and EM bilinguals (-0.001ms). There was also a significant condition difference in the VOT production of the voiceless bilabial /p/, $F(3, 25) = 49.83$, $p < .001$, as it was produced with short lag by the MGs (15ms) and GM bilinguals (12ms), but with long lag by the AusEs (77ms) and EM bilinguals (76ms). Similarly, significant differences were found in productions of the voiced coronal /d/, $F(3, 25) = 75.666$, $p < .001$, produced with voicing lead by the MGs (-133ms) and GM bilinguals (-123ms) but with short lag by the AusEs (5ms) and EM bilinguals (9ms). There was also a significant condition difference in the VOT production of the voiceless coronal /t/, $F(3, 25) = 49.831$, $p < .001$, as it was produced with short lag by the MGs (17ms) and GM bilinguals (19ms), but with long lag by the AusEs (92ms) and EM bilinguals (82ms). Post hoc tests revealed that EM bilinguals produced both voiced and voiceless AusE stop consonants with significantly longer VOT values than did GM bilinguals, consistent with VOT differences between AusE and MG. They matched the VOT of the AusE monolinguals for the voiceless stops /p, t/ and the voiced stops /b, d/. GM bilinguals produced VOTs consistent with MG monolinguals for both the voiceless stops /p, t/ and the voiced stops /b, d/. This pattern of language of presentation differences in bilingual VOT production, and their similarity to the corresponding values for monolinguals, was upheld for all stops across positional contexts.

The present findings support the existence of abstract phonological voicing contrasts in both languages, as the speakers reliably produced different VOTs for the voiced and voiceless stops. The results also support the existence of language-specific phonetic categories, as the two bilingual groups produced clearly different VOT values for MG versus AusE stops. Existing L2 speech production theories (e.g. SLM) must be modified to account for this bilingual ability to shift language modes not only at the level of vocabulary and grammar, but also at the level of fine-grained speech details. In fact, to our knowledge, no current theory of speech production provides an explanation for systematic variation in contrastive phonetic settings depending on linguistic context.

Future research will examine whether switching the linguistic context for our two groups of bilinguals in a second recording session will result in language-relevant shifts in VOT within subjects, consistent with those observed between subjects in the present study. Change in language use and language setting, as that which occurs when living overseas, has been shown to change a bilingual's production of both L1 and L2 stop consonants (Sancier & Fowler, 1997). It is predicted that similar shifts in production will be observed if the linguistic context *within* the laboratory is carefully controlled and manipulated relatively naturalistically.

References

Caramazza, A., Yeni-Komshian, G. H., Zurif, E. B., & Carbone, E. (1973). The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals. *Journal of the Acoustical Society of America*, 54, 421-428.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research* (pp. 233-277). Timonium MD: York Press.

Flege, J.E., & Eefting, W. (1987). Cross-language switching in stop consonant perception and production by Dutch speakers of English. *Speech Communication*, 6, 185-202.

Green, D. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, 1, 67-81.

Grosjean, F. (2001). The bilingual's language modes. In J. L. Nicol & T. D. Langendoen (Eds.), *Language processing in the bilingual*. Oxford: Blackwell.

Hazan, V. L., & Boulakia, G. (1993). Perception and production of a voicing contrast by French-English bilinguals. *Language and Speech*, 36, 17-38.

Magloire, J., & Green, K. P. (1999). A cross-language comparison of speaking rate effects on the production of voice onset time in English and Spanish. *Phonetica*, 56, 158-185.

Sancier, M. L., & Fowler, C. A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics*, 25, 421-436.

Accessing psycho-acoustic perception and language-specific perception with speech sounds

Molly Babel & Keith Johnson

Department of Linguistics, University of California, Berkeley
mbabel@berkeley.edu; keithjohnson@berkeley.edu

In this paper, we reexamine the results of four experiments that explore the difference between language-specific perception and psycho-acoustic auditory perception of speech sounds. These experiments examine this processing difference through two types of tasks: speeded discrimination and similarity rating. These tasks are preferred over labeling tasks as they allow listeners to judge the perceptual similarity of two tokens without having to make explicit judgments about the proper orthographic representation for the sound or whether such a symbol even exists within their repertoire (cf. Best et al. 2001). For the similarity rating task, listeners are asked to compare two stimuli that are presented with a 100 ms ISI on a 5-point equal interval scale ranging from “very similar” to “very different”. This task allows for more leisurely processing of the tokens and, therefore, language-specific perception is evaluated. For the speeded AX discrimination tasks, two stimuli are presented consecutively with a 100 ms ISI. Listeners are instructed to press assigned buttons to indicate if the two sounds are “same” or “different”. Listeners are instructed to respond within 500 ms; their response time and accuracy are presented on the screen between trials to motivate their performance. The fast-paced nature of this task is designed to bypass linguistic processing and hone in on pure acoustic similarity (Pisoni, 1973; Pisoni & Tash, 1974; Fox, 1984).

The first set of experiments (Johnson & Babel, 2007) investigated the perception of voiceless fricatives [f θ s ʃ x h] in three vowel environments [i a u] in V_1CV_1 sequences by Dutch and English listeners using the two tasks described above. In the similarity rating task, the listener rating scores were analyzed in a repeated measures ANOVA. There was a significant main effect of vowel context ($F[2,52] = 65.3$, $p < 0.01$) and consonant pair ($F[14, 364] = 94.5$, $p < 0.01$). The consonant pair and vowel interaction was significant ($F[28, 728] = 15.5$, $p < 0.01$). Crucially, a consonant pair by language interaction was also found ($F[14,364]=3.8$, $p < 0.01$), along with a three-way consonant pair by vowel by language interaction ($F[28,728]=1.6$, $p < 0.05$). For the speeded discrimination task only correct “different” responses were analyzed in a repeated measures ANOVA. There were no significant language effects. A main effect of consonant pair ($F[14, 393] = 15.8$, $p < 0.001$) and vowel environment ($F[2, 11] = 18.9$, $p < 0.001$) were found. There was also an interaction between consonant pair and vowel ($F[28, 888] = 18.8$, $p < 0.001$).

In the second set of experiments (Babel & Johnson, 2007), the perception of Russian palatalization was compared across Russian and English listeners. Four degrees of Russian palatalization taken from Russian words were implemented in the task design: CV [mat] ‘checkmate’, C^jV [mʲat] ‘rumpled’, C^jjV [sudʲja] ‘judge’, and C^jijV [zmʲiju] ‘snake’. The stimuli were formatted as one of the four aforementioned sequences with a vowel /a u i/ and one of six possible onsets: /m v b d l r/. In the similarity rating task, a repeated measures ANOVA found main effects for language ($F[1, 21]=9$, $p < 0.01$), palatalization pair ($F[5, 105]=369.2$, $p < 0.001$) and vowel context ($F[2, 42]=21$, $p < 0.001$). There were also significant interactions between palatalization pair and vowel ($F[10, 210]=18$, $p < 0.001$) and palatalization and language ($F[5, 105]=3.4$, $p < 0.01$). Again, “same” pairs of stimuli and incorrect responses were removed prior to the analysis for the speeded discrimination task. A repeated measures ANOVA revealed a main effect of palatalization pair ($F[5, 16] = 257.28$, $p < 0.001$) and consonant ($F[5, 4] = 13.63$, $p < 0.05$). There were significant interactions between language and palatalization ($F[5,16] = 4.52$, $p < 0.01$), palatalization and consonant ($F[25, 460] = 2.28$, $p < 0.001$), and palatalization and vowel ($F[10, 230]=6.83$, $p < 0.001$). The ANOVA also returned a three-way interaction between palatalization, consonant, and vowel ($F[50, 1105] = 2.7$, $p < 0.001$). Planned comparisons of the language and palatalization interaction revealed that there were no significant differences between listener groups with the palatalization pairs.

In both sets of experiments, the similarity rating tasks found strong language effects that reflect the phonological organization of the languages. For example, Dutch listeners do not contrast /s/ and /ʃ/ and they rated these sounds as more similar-sounding than English listeners in the rating tasks. The speeded AX discrimination experiments did not reveal language effects; all listeners in these tasks performed comparably, regardless of language background. The results of these experiments suggest that psycho-acoustic perception can be evaluated apart from linguistic perception.

Other work using this paradigm found language effects in AX discrimination as well as rating tasks (Boomershine, Hall, Hume, & Johnson, in press; McGuire, 2007). To reconcile these results with our work, we adopted a binned reaction time analysis with the Boomershine et al. data and the McGuire data. The analysis examined the language effects in the AX discrimination tasks in their experiments through three binned range groups. Range 1 was composed of responses made faster than 500 ms, Range 2 were those responses logged between 500 and 800 ms, and Range 3 were those made after 800 ms. It was predicted that language effects would be found only in slower responses that allowed for linguistic processing. This analysis and its predictions follow from Fox (1984). Fox demonstrated that the lexical effects of the Ganong effect disappear in responses made prior to 800 ms. Our analysis found the language effect in the McGuire data was significant only in the response latencies greater than 800 ms. In the binned range analysis of the Boomershine et al. data the language effects remained, but were the result of consistently longer response latencies by the Spanish listeners and not in any linguistically interpretable way. We argue that the language effect in their data is retained due to delayed reaction times and the bilingualism of the Spanish participants. The work reported argues for multiple stages in perception. Specifically, we provide evidence that it is possible to experimentally probe a stage of auditory perception that is not affected by language specific perceptual warping.

Acknowledgements

This research has been supported by NIH grant R01 DC004421-05 awarded to the second author. Thanks to Grant McGuire, Amanda Boomershine, Kathleen Currie Hall, and Elizabeth Hume for sharing their raw data with us.

References

Babel, M. & Johnson, K. (2007) Cross-linguistic Differences in the Perception of Palatalization. *Proceedings of the International Congress of the Phonetic Sciences*.

Best C.T., McRoberts G.W., & Goodell E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America*, 109 (2), 775-794.

Boomershine, A., Hall, K. C., Hume, E., & Johnson, K. (in press). The impact of allophony vs. contrast on speech perception. In P. Avery, E. Dresher & K. Rice (Eds.), *Phonological Contrast: Perception and Acquisition*. New York: Mouton de Gruyter.

Fox, R.A. (1984). Effect of lexical status on phonetic categorization. *Journal of Experimental Psychology: Human Performance*, 10, 526-540.

Johnson, K. & Babel, M. (2007) Perception of fricatives by Dutch and English speakers. *UC Berkeley Phonology Lab Annual Report 2007*, 302-325.

McGuire, G. (2007). *Phonetic category learning*. Ph.D. Dissertation, The Ohio State University.

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics*, 13(2), 253-260.

Pisoni, D. B. & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics*, 15(2), 285-290.

Generalizing over Lexicons to Predict Consonant Mastery

Mary E. Beckman* & Jan Edwards#

*Department of Linguistics, Ohio State University, USA; mbeckman@ling.osu.edu
#Department of Communicative Disorders, University of Wisconsin - Madison, USA; jedwards2@wisc.edu

When children first begin to produce vocalizations that listeners recognize as meaningful words of the ambient language, they show characteristic consonant errors. These misarticulations are generally perceived in terms of categorical processes that delete target consonants or that substitute one adult target consonant for another. For example, English-acquiring children are stereotypically perceived as substituting alveolar stops for velars and voiced stops for aspirated, so that a 2-year-old's productions of *goat* and of *tub* both might be transcribed as having an initial /d/ (or [t] in a more narrow "phonetic" transcription). There are two seemingly contradictory characterizations of these misarticulations. In work such as Locke's (1983) masterful review monograph, we see broad generalizations about typical consonant inventories and typical misarticulations, which seem to support Jakobson's (1941) "laws of irreversible solidarity" — i.e., "implicational universals" such that mastery of more "marked" sounds implies the prior mastery of less "marked" sounds. By contrast, the detailed quantitative observations in longitudinal cross-language studies such as Vihman (1993) show an enormous variability in production patterns in young children, and at least part of this variability seems to be correlated with differences in consonant frequencies across the different lexicons.

In this paper, we examine whether these two characterizations can be reconciled by relating them to a broad-grained typological generalization about consonant inventories. Adopting Lindblom and Maddieson's (1988) terms, the generalization is that smaller inventories tend to include only simple "basic" consonants that also occur in languages with "elaborated" or "complex" ones. For example, 413 of the 451 languages in the UPSID database (Maddieson & Precoda, 1989) are listed as having [t] whereas only 111 are listed as having the aspirated counterpart [t^h]. Similarly, 403 languages are listed as having [k], whereas only 13 are listed as contrasting the palatalized [k^j] and only 3 have the "complex" stop [k^{jh}]. This generalization echoes the implicational universals that seem to hold in broad-grained characterizations of the relative order of consonant mastery across languages. That is, just as languages tend to incorporate consonants such as [t] and [k] before consonants such as [t^h], [k^j], and especially [k^{jh}] children tend to master the basic consonants first.

Of course, children do not master consonants in isolation. Indeed, many of the most basic consonants cannot be produced audibly in isolation. The "sound" of [t] or [k] when there is no neighboring vowel or vowel-like release is a silence indistinguishable from the "sound" of [t^h] or [k^j]. Mastery of a consonant implies mastery of a set of words containing the consonant in a particular set of contexts that allow the consonant to be audibly pronounced for the child to hear and reproduce. A first question we ask, then, is whether the qualitative generalization about consonant inventories across languages also holds as a set of quantitative generalizations about the inventories of words in specific languages. Do basic consonants which occur in many languages also tend to occur in many words in the lexicons of these languages? Conversely, do elaborated and complex consonants that occur in few languages occur in few words in the lexicons of those languages? If the broad-grained qualitative generalization about consonant inventories also holds as a fine-grained quantitative generalization about lexicons, then this might help explain the developmental universals. That is, since mastering a consonant means learning to perceive and reproduce it in a particular set of words, we might expect a very young child to be more accurate in producing consonants that occur in many of the words of the ambient language and less accurate in producing ones that occur in few words. Insofar as these frequencies are similar across languages because of the forces of Evolutionary Phonology (Blevins, 2006), we might expect similarities to the accuracy rates and error patterns of children.

To address these issues, we compared accuracy of target consonants in a database of productions elicited in a word-repetition task from 2- and 3-year-old children acquiring Cantonese, English, Greek, or Japanese and we also compared relative frequencies of the same target consonants in the lexicons of these languages. For each language, the target consonants were all of the lingual obstruents that occur word-initially in at least three words that a young child is likely to know. For example, in Japanese, target [t] was elicited in *tamago* ‘egg’, *tebukuro* ‘glove’, *tisshu* ‘tissue’, *tora* ‘tiger’, and 8 other words. We are interested in lingual obstruents because these are less “transparent” to the young language learner than either labials (where there are clear visual clues to the place and manner of articulation) or glides (where the auditory feedback about lingual posture is available simultaneously with the somatosensory feedback). To master the articulation of the [t] in both *tebukuro* and *tora*, for example, a Japanese-acquiring child must deduce not only that the silence at the beginning of each word corresponds to an interval when the tongue blade is pressed against the upper incisors to seal off the oral cavity, but also that the posture of the body of the tongue behind the seal is responsible for the difference in second formant transition after its release.

Accuracy rates for “shared” sounds were more or less correlated across language pairs (r^2 ranged from 82% for Cantonese/Greek to 2% for Japanese-English), but pair-wise correlations of frequencies missed many sounds. We therefore developed a more global measure of expected frequency by first ranking frequencies of all of the consonants that occurred in any language, counting all unattested consonants for a language as equally at the bottom of the list. We then correlated the log frequencies of the consonants within each language against the mean ranks for the three other languages. These mean ranks accounted for at least 20% of the variance in the log frequencies of each language’s inventory of lingual obstruents, although the correlation was significantly predictive only for Greek and English. We interpret this result as offering modest support to the idea that the same consonants tend to occur in many words across languages and that “marked” consonants that are attested in few languages tend to occur in fewer words in those languages.

To test the corollary idea that children are quicker to master consonants that occur in many words, we regressed the accuracy rates for the consonants in the word productions elicited from the twenty children for each language against both the log frequencies of the consonants in the target language and their mean ranks across the languages. The different components of the regression models accounted for different amounts of the variation in accuracy — e.g., in Cantonese, the global frequency ranking was the only significant component of the model, whereas in English only the language-specific consonant frequency was significant. We interpret these results in terms of historical contingencies that have conspired to make for some unexpected patterns of lexical frequencies. For example, voiced fricatives, which are rare cross-linguistically, developed historically in Greek from voiced stops, and are much less accurate than predicted from their high frequencies relative to the voiced stops that occur in recent loanwords such as γκολ [gol] ‘goal’. Also, in both Cantonese and English, the aspirated stops are less accurate than the unaspirated stops. In Cantonese, the aspirated stops are also less frequent than the unaspirated stops, since the latter represent two Middle Chinese types (voiced and voiceless unaspirated) which merged historically. In English, by contrast, the aspirated stops are more frequent, having developed historically from the “basic” voiceless stops of common Germanic. In short, calling two sounds the “same” across different languages is an identity by analogy, not homology. Each child acquires a phonology by learning the auditory and articulatory patterns of the words of the specific ambient language. Different histories for analogous consonants can lead to different frequencies, different accuracies, and different relationships between frequency and accuracy.

Acknowledgements

Work supported by NIDCD grant 02932 to Jan Edwards.

References

Blevins, J. (2006). A theoretical synopsis of Evolutionary Phonology. *Theoretical Linguistics*, 32 (2), 117-166.

Jakobson, R. (1941). *Kindersprache, Aphasie und allgemeine Lautgesetze*. Uppsala: Almqvist & Wiksell.

Lindblom, B., & Maddieson, I. (1988). Phonetic universals in consonant systems. In L. M. Hyman, & C. N. Li (Eds.), *Language, Speech, and Mind* (pp. 62-78). New York: Routledge.

Locke, J. L. (1983). *Phonological acquisition and change*. New York: Academic Press.

Maddieson, I., & Precoda, K. (1989). Updating UPSID. *Journal of the Acoustical Society of America*, 86(S1), S19. (UPSID-PC downloaded from http://www.linguistics.ucla.edu/faciliti/sales/software.htm)

Vihman, M. M. (1993). Variable paths to early word production. *Journal of Phonetics,* 21, 61-82.

Identity Avoidance Between Non-Adjacent Consonants in Artificial Language Segmentation

Natalie Boll-Avetisyan & René Kager

Utrecht Institute of Linguistics, Utrecht University, Netherlands; natalie.boll-avetisyan@let.uu.nl

Phonotactic distributions over lexicons correlate with human performance (Coleman & Pierrehumbert, 1997; Vitevitch & Luce, 1999; Pitt & McQueen, 1998). Although it is known that phonotactic probability is independent of lexical analogy (Bailey & Hahn, 2001), the issue remains whether gradient distributions over the lexicon are represented in the grammar in the form of abstract constraints involving natural classes (Frisch, Pierrehumbert, & Broe, 2004; Hayes & Wilson, 2007).

This paper supports the hypothesis that gradient distributions over the lexicon may be represented as abstract constraints, which affect segmentation. We tested this hypothesis using artificial language segmentation, a task which is known to be affected by phonological properties of the native language (Onnis, Monaghan, Chater, & Richmond, 2005). We focussed on non-adjacent C_C dependencies, allowing us to test effects of abstract constraints while controlling for phonotactic probabilities of the stimulus material. It has been found that statistical calculations over triphones (e.g. CVC) are uneffective for non-word processing (Bailey & Hahn, 2001). Yet artificial language learning studies show that statistical dependencies between non-adjacent consonants are learnable (Newport & Aslin, 2004). Many languages restrict the co-occurrence of homorganic consonants across intervening vowels (e.g. Frisch & Zawaydeh 2001). Artificial language studies and cross-linguistic studies are consistent with the prediction that constraints on non-adjacent consonants with a probabilistic basis in the lexicon affect artificial language segmentation.

Lexical statistics

We collected cooccurrence statistics on non-adjacent consonants in Dutch using Observed/Expected values, with $E(C_1VC_2)$ calculated as $p(C_1) * p(C_2) * N_{CVC}$ in CELEX (Baaijen, Piepenbrock, & Gulikers 1995). Two observations stood out. Pairs of labials (P={p b f v m w}) are strongly under-represented (O/E = 0.45) as compared to coronal pairs (T={t d s z n l r}; O/E = 0.77). In word-initial position (where $E = p(C_1) * N_{\#C}$), labials are strongly over-represented (O/E = 2.0), whereas coronals are only slightly under-represented (O/E = 0.60). From this distribution, we hypothesized two abstract constraints. OCP-LAB forbids pairs of consecutive labials. ALIGN-LAB requires labials to occur word-initially. In order to examine whether these constraints are used for segmentation, we ran two artificial language experiments.

Experiment 1

Native speakers of Dutch were trained in an artificial language that consisted of six P syllables (P_1={po, be, ma}, P_2={pa, bi mo}) and three T syllables (T={tu, do, ne}), which were concatenated into a synthetic speech stream (…$P_1P_2TP_1P_2TP_1P_2T$…) with flat prosody. Other potential cues for segmentation from lexical statistics, such as transitional probabilities between syllables and segments, positional syllable frequency, and the distribution of nasality and voicing were controlled. If participants were able to segment words from this stream, predictions from OCP-LAB and ALIGN-LAB were that that PTP words should come out best, and TPP words should come out worst.

In the test phase, participants had to decide which of two CVCVCV strings was a word from the language they had been trained on. Participants (N=42) were assigned to one of three test conditions: PTP-PPT, PTP-TPP, or PPT-TPP. The 48 test pairs were matched for lexical factors (cohort density, lexical neighbourhood density). As predicted, PTP words, which satisfy OCP-LAB and ALIGN-LAB, were preferred over PPT ($p < .001$) and TP ($p < .05$) to be words of the artificial language in which participants had been trained (see Figure 1). In particular, the strong preference PTP > PPT suggests that OCP-LAB has an impact on segmentation. Results of the individual test items did not correlate with O/E values for consonant combinations, which supports our hypothesis that segmentation preferences reflect an abstract constraint, rather than experience with individual consonant combinations. In the PPT-TPP comparison, the predicted preference for PPT, due to ALIGN-LAB alone, was not found. This suggests that the PTP segmentation may have been too dominant over the PPT or TPP segmentations for a PPT-TPP comparison to be possible at all. A replication of the experiment using a different syllable set yielded the same result.

Experiment 2

In order to rule out a possible interpretation that the result was not caused by OCP-LAB, but by a general perceptual preference for identity at edges (ABA > AAB, ABB, regardless of values of A,B), we ran a second experiment. Participants were trained in an artificial language similar to the language in Experiment 1, the only difference being that P_1 was replaced by T (T_2={ta, di, no}). In Dutch, consecutive coronals are less under-represented than labials. Accordingly, we hypothesized that any co-occurrence effects will be over-ruled by ALIGN-LAB, and expected a PTT segmentation.

As in Experiment 1, we assigned participants (N=42) to one of three test conditions, constrasting TPT-TTP, TPT-PTT, or PTT-TTP pairs. As predicted, PTT words were favoured over TPT ($p < .001$) and TTP ($p < .001$). In the TTP-TPT comparison, none of the words was favoured, which supports the prediction that ALIGN-LAB affected segmentation, not a general preference for ABA.

Conclusion

The segmentation results suggest that probabilistic distributions over the lexicon may be represented as abstract constraints in the grammar. The hypothesized constraints were abstract since their effects could not be reduced to low-level statistical properties of the stimulus material. Future studies may explore the consequences of our findings for models of speech segmentation.

References

Baaijen, R., Piepenbrock, R. & Gulikers, L. (1995). *The CELEX Lexical Database* (CD-ROM). Linguistic Data Consortium, University of Pennsylvania, Philadelphia, PA.

Bailey, T., & Hahn, U. (2001). Determinants of wordlikeness: phonotactics or lexical neighborhoods. *Journal of Memory and Language* 44, 568–591.

Coleman, J., & Pierrehumbert, J. (1997). Stochastic phonological grammars and acceptability. *3rd Meeting of the ACL Special Interest Group in Computational Phonology*. ACL: Somerset NJ. 49-56.

Frisch, S., & Zawaydeh, B. (2001). The psychological reality of OCP-Place in Arabic. *Language,* 77, 91–106.

Frisch, S., Large, N., & Pisoni, D. (2000). Perception of wordlikeness: Effects of segment probability and length on the processing of nonwords. *Journal of Memory and Language*, 42, 481-496.

Frisch, S., Pierrehumbert, J., & Broe, M. (2004). Similarity avoidance and the OCP. *Natural Language & Linguistic Theory*, 22, 179-228.

Hayes, B., & Wilson, C. (2007). A maximum entropy model of phonotactics and phonotactic learning. To appear in *Linguistic Inquiry*.

Newport, E., & Aslin, R. (2004). Learning at a distance: I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, 48, 127-162.

Onnis, L., Monaghan, P., Richmond, K., & Chater, N. (2005). Phonology impacts segmentation in online speech processing. *Journal of Memory and Language*, 53, 225-237.

Pitt, M., & McQueen, J. (1998). Is compensation for coarticulation mediated by the lexicon? *Journal of Memory and Language*, 39, 347-370.

Vitevitch, M., & Luce, P. (1999): Probabilistic phonotactics and neighbourhood activation in spoken word recognition. *Journal of Memory and Language*, 40, 374-408.

Weber, A., & Cutler, A. (2006). First-language phonotactics in second-language listening. *Journal of the Acoustical Society of America*, 119, 597-607.

The Perception of a Vowel Contrast by Bilingual Children

Ivana Brasileiro

Utrecht Institute of Linguistics OTS, Utrecht University, The Netherlands; Ivana.Brasileiro@let.uu.nl

In this study I deal with the acquisition of a Dutch vowel contrast by children being raised bilingually in Dutch and Brazilian Portuguese (BP). The main question I address is the possible effect of simultaneous bilingualism in the phonetic detail of children's categorization. By simultaneous bilingual children I mean children who have been exposed to both of their languages from birth. The contrast tested was the Dutch /aː/-/ɑ/. Crucially, this Dutch contrast involves spectral and durational differences and is not found phonemically in BP.

One hundred and sixteen participants took part in this study.

1) 29 bilingual children in Dutch and Brazilian Portuguese (3;5 – 7;1 years old) raised in The Netherlands;
2) 19 bilingual children in Brazilian Portuguese and Dutch (4;7 – 7;1 years old) raised in Brazil;
3) 29 Dutch monolingual children (3;9 - 6;5 years old);
4) 14 Dutch monolingual adults (19 - 35 years old);
5) 14 Brazilian Portuguese monolingual children (4;6 - 6;7 years old);
6) 13 Brazilian Portuguese monolingual adults (23 - 27 years old.)

There are two groups of bilinguals involved in this research, namely one raised in The Netherlands (group 1) and one raised in Brazil (group 2). This separation allows us to control for language dominance.

The stimuli used were 12 synthesized vowels, manipulated in two dimensions, a spectral one (F1 and F2) and a durational one. Each dimension was manipulated in four steps, creating a 4 x 4 matrix. For this experiment only the edges were used as we were trying to reduce the length of the experiment. The vowels are synthesized in such a way as to range from the Dutch /ɑ/ to the Dutch /aː/ using some of the acoustic space reserved for the BP /a/, as shown in Figure 1.

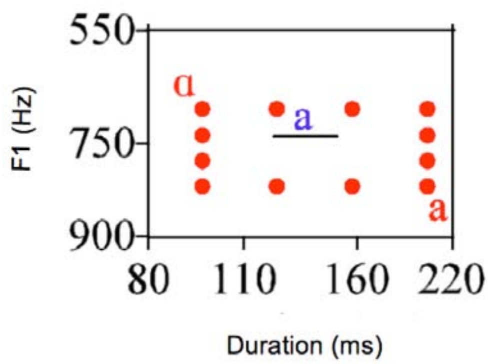


Figure 1: The 12 synthetic vowels used in this experiment in comparison to the Dutch and BP low vowels (BP vowel is underlined)

All groups were tested in an XAB experiment where stimuli were presented through headphones. Participants heard three tokens in a row (X, A, and B) and had to decide whether the first token (X) matched the second (A) or the third sound (B). X was any of the 12 tokens synthesized vowels. A and B were each of the end points of the continuum, i.e. one of the two Dutch vowels, either /aː/ or /ɑ/. In order to make the experiment accessible for the children, it has been embedded into a computer game. Most children reported to find it easy and fun.

To calculate participants' perceptual reliance of spectrum and duration, I followed the same procedure used in Flege, Bohn & Jang (1997) and Escudero & Boersma (2004). We calculated the reliance of spectral differences to perceive the vowel contrast, i.e. *spectrum reliance*, by subtracting the number of /ɑ/ responses of the top horizontal edge of the continuum from the scores of the bottom horizontal edge (see figure 1 for reference). The value was then averaged across the number of tokens in each row (4 in this case). Similarly, the *duration reliance* was computed by subtracting the amount of /ɑ/ responses of the right vertical edge of the continuum from the scores of the left vertical edge. This value was also averaged across the number of tokens in each column (4).

Although data analysis is still in progress, our preliminary results (over 75% of the data) show no significant difference between monolingual and bilingual children's perceptual responses. A regression analysis using the dummy variables "bilingualism", "childhood" and "place where raised" reveals that "childhood" is the only significant factor. This indicates that bilingualism does not affect the categorization of sounds and that bilinguals are able to form native-like categories even in their non-dominant language. These results also show that up to the age tested, i.e. up to about 5 years of age, children, regardless if they are monolingual or bilingual, have not yet acquired the adult cue-weighting norm. Furthermore, spectrum reliance correlates significantly with age and vocabulary size. Bosch & Sebastián-Gallés (2003) have suggested that infants exposed to two languages have some delay in discriminating native sounds when compared to infants exposed to one language only. Sundara, Polka & Genesee (2006) found similar results with pre-school aged children. The results of the current study, however, are not in line with these conclusions, as we find bilingual children patterning with monolingual children.

Additional interesting results come from the cross-linguistic perception of the on the Dutch contrast by the BP monolingual listeners. These results suggest that two distinct categorizations strategies might be taking place (as previously claimed by Brasileiro & Escudero, 2006). Specifically, some BP monolingual listeners seem to be using the durational dimension to categorize the Dutch vowels (/aː/ and /ɑ/) whereas in BP duration is not used phonemically at the segmental level. I interpret these results suggesting that these listeners are categorizing both Dutch vowel onto the same BP category (/a/), and start using duration as an extra linguistic strategy in order to deal with the somewhat ambiguous input. On the other hand another group of BP monolingual listeners uses a lot of spectrum, more than the native Dutch listeners. I suggest that these listeners are categorizing the Dutch /aː/ onto BP /a/ and the Dutch /ɑ/ onto the BP /ɔ/. In this case, the contrast can be made purely on the basis of spectrum.

References

Brasileiro, I., & Escudero, P. (2005). A three-way comparison in perceptual development: Monolingual children vs. simultaneous bilingual children vs. adult L2 learners. Paper presented at *EuroSLA 16*, Antalya: Turkey.

Bosch, L., & Sebastián-Gallés, N. (2003). Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life. *Language and Speech,* 46, 217-243.

Escudero, P., & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition,* 26, 551- 585.

Flege, J. E., Bohn, O.-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics,* 25, 437-470.

Sundara, M., Polka, L., & Genesee, F. (2006). Language-experience facilitates discrimination of d-ð in monolingual and bilingual acquisition of English. *Cognition,* 100, 369-388.

How Acoustically Reduced Forms Activate the Lexicon: Evidence from Eye-tracking

Susanne Brouwer[1], Holger Mitterer[2], & Mirjam Ernestus[3]

[1]Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands, susanne.brouwer@mpi.nl
[2]Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands, holger.mitterer@mpi.nl
[3]Radboud University & Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands, mirjam.ernestus@mpi.nl

Most research on spoken word comprehension has focused on carefully articulated speech that is read aloud by selected speakers (Cutler, 1998). But the type of speech we most often encounter is spontaneous speech, in which no attention is paid to careful pronunciation. The production of a word shorter than its citation form is called reduction, which is highly frequent in casual speech (Ernestus, 2000; Johnson, 2004). The challenge for models of word comprehension is to explain how listeners recognize reduced forms such as [pjutər] which deviate drastically from their canonical counterpart [kɔmpjutər] 'computer'.

Several mechanisms have been proposed to explain the recognition of reduced forms. All reduced forms of a word may have their own lexical representations and recognition just consists of mapping the form on the right representation ([pjutər]) mapped on /pjutər/, Johnson, 2004). In contrast, every word may be lexically represented by just one form, indicating the word's citation form. The reduced form may then be reconstructed to the word's citation form at the pre-lexical level on the basis of fine phonetic detail in the signal (Mitterer & Blomert, 2003). In this study, we investigated the mechanisms involved in the recognition of reduced forms.

Crucially, the two accounts differ in their prediction which lexical candidates act as serious competitors for reduced forms in the recognition process. The exemplar-based account predicts competition especially from words sharing their pronunciation with the reduced form (e.g., [pjupɪl] 'pupil' for [pjutər]). The reconstruction account predicts that the most important competitors are those that are phonologically similar to the word's citation form (e.g., [kɔmpɛnjən] 'companion' for [kɔmpjutər]). We investigated the relevance of the different competitors using the visual-world paradigm (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995), which measures participants' eye movements to printed words closely time-locked to the unfolding speech signal. This eye-tracking paradigm provides a sensitive, continuous measure of competition effects in spoken word recognition.

We created two types of competitors: "underlying" competitors which were similar in pronunciation to the word's citation form and "surface" competitors which were similar to the pronunciation of the reduced form of the target word. For the experimental trials, we displayed an underlying competitor (*competent*, /kɔmpətɛnt/), a surface competitor (*poetsen* 'to clean', /putsə/) and two phonologically unrelated distractors (*juweel* 'jewel', *vakantie* 'holiday') on the computer screen, while participants listened to spontaneous utterances from the Spoken Dutch Corpus (Oostdijk, 2000) containing reduced ([pjutər]) and non-reduced forms ([kɔmpjutər]). Thus, for the critical trials, the target word was never present on the screen. Subsequently, the participants' task was to click in the center of the screen. For the filler trials, one of the written words was present in the utterance. Participants were then asked to click on the word they heard in that utterance.

Figure 1 presents the fixation proportions from target onset (0 ms) to 1400 ms thereafter for the (A) Non-reduced and (B) Reduced forms of the target words. The results for the non-reduced forms show, unsurprisingly, that the underlying competitor, which is most similar to the presented pronunciation of the target word, is overall significantly more activated than the surface competitor. More importantly, during the presentation of the reduced forms, the surface competitors were significantly more strongly activated in the early time window (400-800 ms) than the underlying competitors. This activation reflects an inability to immediately reconstruct the reduced forms to their canonical forms. That is, information in the fine phonetic detail is not strong enough to block the activation of the surface competitor. In the late time window (800-1200 ms), we observe a strong rise of the underlying competitor. This indicates that restoration of the reduced form based on phonetic detail might occur, but that the process is time-consuming.

In conclusion, the clear activation of both competitors suggests that the comprehension process involves the activation of both the citation form and reduced representations, whereas previous studies have only been able to document a role for the citation form (e.g., Ernestus, Baayen & Schreuder, 2002). This study is the first to show that surface competitors also contribute to the recognition process.

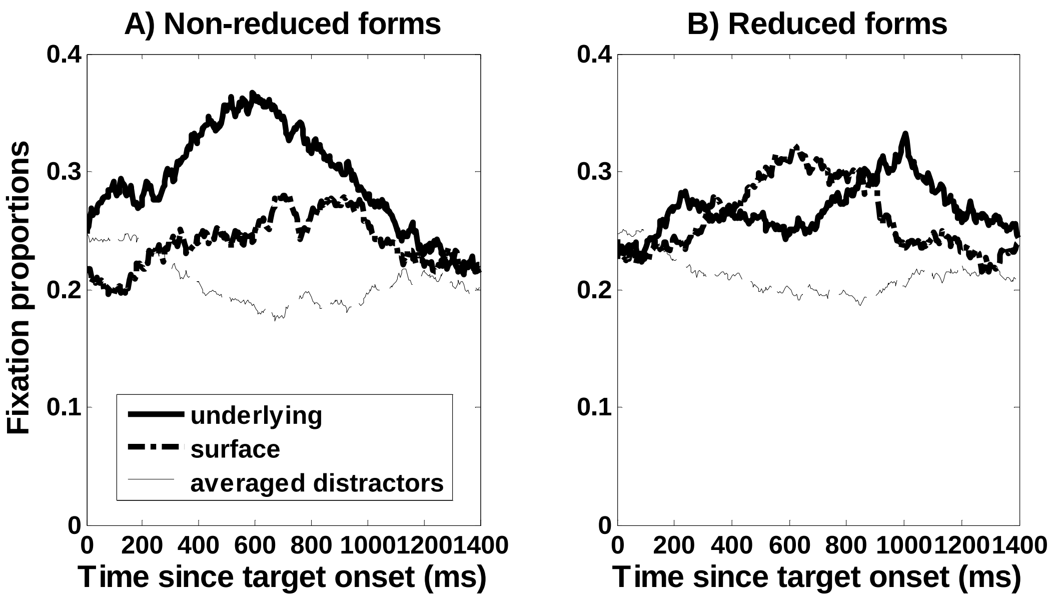


Figure 1: Fixation proportions to the underlying competitor, the surface competitor and the two averaged distractors for (A) Non-reduced forms and (B) Reduced forms.

References

Cutler, A. (1998). The recognition of spoken words with variable representation. In Danielle Duez (Ed.) *SPoSS - Sound Patterns of Spontaneous Speech* (pp. 83-92). Aix-en-Provence, France.

Ernestus, M. (2000). *Voice assimilation and segment reduction in casual Dutch. A corpus-based study of the phonology-phonetics interface.* Utrecht: LOT.

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 K. Yoneyama & K. Maekawa (Eds.). *Spontaneous Speech: Data and Analysis. Proceedings of the 1st session of the 10th International Symposium* (pp. 29-54). Tokyo, Japan.

Mitterer, H., & Blomert, L. (2003). Coping with phonological assimilation in speech perception: Evidence for early compensation. *Perception and Psychophysics,* 65(6), 956-969.

Oostdijk, N. (2000). The Spoken Dutch Corpus project. *The ELRA Newsletter*, 5, 4-8.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268, 1632-1634.

Gradient phonotactics in the Gitksan lexicon

Jason Brown* & Gunnar Ólafur Hansson#

*Department of Linguistics, University of British Columbia, Canada; jcb@interchange.ubc.ca
#Department of Linguistics, University of British Columbia, Canada; ghansson@interchange.ubc.ca

Patterns in the distribution of consonants in the lexicon, and patterns of consonant *co-occurrence* in particular, have in recent years been subjected to careful study in a number of languages, e.g., English (Berkeley, 2000), Arabic and other Semitic languages (Frisch, Pierrehumbert & Broe, 2004), and Muna (Coetzee & Pater, to appear). These studies have focused in particular on dissimilatory consonant co-occurrence restrictions (Obligatory Contour Principle effects), and the extent to which gradient patterns of relative under- vs. overrepresentation of specific consonant co-occurrences is correlated with the relative similarity of the segments in question. The results of such studies have important implications for our view of the relationship between the lexicon and the phonological grammar and between gradient and categorical sound patterns. Questions also arise in relation to the similarity metric, such as to what extent it is influenced by inventory structure, and whether it can be subject to language-particular weighting of distinctive features.

The indigenous languages of the Pacific Northwest are interesting from this perspective, as they often have consonant inventories which differ in many respects from the kinds of languages in which co-occurrence patterns have been explored in detail. In this study, we examine consonant co-occurrences in the lexicon of Gitksan, an endangered Tsimshianic language of British Columbia (Canada), whose inventory displays various features typical of the Northwest Coast linguistic area. For example, we find a relatively large class of dorsals (uvulars and velars, the latter exhibiting a labialization contrast), the set of coronals includes lateral obstruents (/tɬ’, ɬ/), and the main laryngeal contrast is one of glottalization: pulmonic vs. ejective plosives (/q/ : /q’/) and pulmonic vs. glottalized sonorants (/n/ : /n’/).

Methods

The Gitksan data we analyzed consist of all root entries from the electronic lexical database of Brown (forthcoming), which in turn draws on a combination of primary fieldwork data and previous published and unpublished sources. Automated searches and calculations were used to obtain Observed/Expected (O/E) ratios for $C_x \ldots C_y$ co-occurrences of various kinds.

Some challenges are presented by the nature of the database and by the phonotactics of Gitksan. At present the database contains only about 645 roots (nominal and verbal roots alike), yielding a total of 1,410 pairs of co-occurring consonants. Given these size limitations, it becomes critical to assess carefully the statistical significance (or lack thereof) of any deviations toward under- or overrepresentation suggested by particular O/E ratios. Furthermore, we are forced to examine co-occurrences not of individual segments, but of aggregate segment classes, and this in turn makes it less straightforward to evaluate the correlation between O/E ratios and relative similarity. In order to address the problem of statistical significance, we followed Martin (2007) in using a Monte Carlo procedure to generate probability distributions from which confidence intervals could be deduced. In each such test, the co-occurrence count for the actual lexicon (the Observed value) was compared against a simulated random-chance distribution, made up of corresponding counts for the 10,000 Monte Carlo lexica (the mean of which constitutes the Expected value), in order to assess whether or not the apparent under- or overrepresentation effect is significant. Figure 1 shows an example of a relatively weak but highly significant underrepresentation effect.

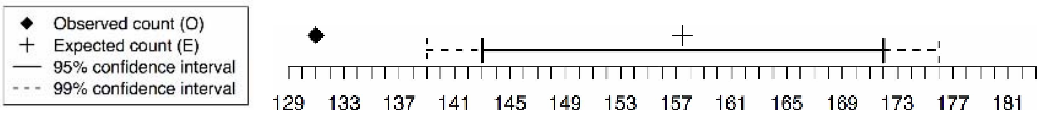


Figure 1: Non-identical Coronal-Coronal CVC pairs (O/E = 0.83; p < 0.001).

Another challenge is that while the canonical root in Gitksan is CVC (comprising 50% of verbal roots in the database), root shapes can be quite diverse: CV, CVCC, CCVC, CVCVC, CCVCC, etc. This motivates a division into 963 *local* $C_x \ldots C_y$ pairs (ones not separated by a C) and 447 *long-distance* $C_x \ldots C \ldots C_y$ pairs; local pairs are further subdivided into 270 *adjacent* CC sequences and 693 *transvocalic* CVC sequences. In order to abstract away from any independent restrictions on clustering, our main focus is on transvocalic (CVC) pairs. Since pairs of identical segments have frequently been found to be exempt from dissimilatory OCP effects (see Frisch et al., 2004), such pairs were generally excluded from consideration.

Results & Discussion

Our main finding is consistent with previous studies on other languages, in that we do find a gradient dissimilatory OCP[Place] effect, which does appear to be sensitive to relative similarity, where that similarity is a reflection of the structure of the inventory. For each major place of articulation, same-place CVC pairs are significantly underrepresented: coronal-coronal pairs (O/E = 0.83), dorsal-dorsal pairs (O/E = 0.54) and labial-labial pairs (O/E = 0.16). As predicted by the natural-classes model of similarity (Frisch et al., 2004), the effect is strongest for labials, which constitute a smaller and less diverse class. The intermediate status of dorsals appears to reflect their relatively large number in the Gitksan inventory, modulated by the fact that, unlike the coronal class, these are all obstruents. Finally, we find small but significant dissimilatory OCP effects over manner categories as well (obstruent vs. sonorant, plosive vs. fricative).

The most notable result, however, concerns the presence of certain gradient *assimilatory* effects along with these general dissimilatory restrictions. For example, although dorsal-dorsal pairs are underrepresented overall, we find that within this class of co-occurrence pairs, “agreeing” ones are strongly overrepresented (O/E = 1.46 for uvular-uvular, O/E = 1.40 for velar-velar), whereas “disagreeing”—and hence less similar—pairs are underrepresented (O/E = 0.57). An analogous situation obtains with glottalization in plosives: among plosive-plosive pairs, which are statistically underrepresented as such, ejective-ejective pairs are significantly overrepresented (O/E = 1.76); so are pulmonic-pulmonic pairs, though the effect is much weaker there (O/E = 1.13). Similarity-based co-occurrence restrictions of an assimilatory kind are a well-known phenomenon (“consonant harmony”; Hansson, 2001; Rose & Walker, 2004), and often take the shape of root-internal co-occurrence restrictions. However, the gradient nature of these effects in Gitksan, and the trade-off between assimilatory and dissimilatory effects, makes this case particularly interesting. In Brown (forthcoming), this interplay is given a formal treatment which combines a correspondence-based analysis of agreement (Hansson, 2001; Rose & Walker, 2004) with an analysis of gradient OCP effects similar to that proposed by Coetzee & Pater (to appear).

Acknowledgements

Thanks to Gitksan language consultants Doreen Jensen and Barbara Sennott, and also to Bruce Rigsby. This research is supported in part by SSHRC Standard Research Grant 410-2004-0710, awarded to Gunnar Ólafur Hansson, and to Jacobs Fund and Phillips Fund grants awarded to Jason Brown.

References

Berkley, D.M. (2000). *Gradient OCP effects*. Ph.D. dissertation, Northwestern University.

Brown, J. (forthcoming). *Theoretical issues in Gitksan phonology*. Ph.D. dissertation, University of British Columbia.

Coetzee, A., & Pater, J. (to appear). Weighted constraints and gradient restrictions on place co-occurrence in Muna and Arabic. *Natural Language and Linguistic Theory*.

Frisch, S.A., Pierrehumbert, J.B., & Broe, M. (2004). Similarity avoidance and the OCP. *Natural Language and Linguistic Theory*, 22, 179–228.

Hansson, G. Ó. (2001). *Theoretical and typological issues in consonant harmony*. Ph. D. dissertation, University of California, Berkeley.

Martin, A. (2007). *The evolving lexicon*. Ph.D. dissertation, University of California, Los Angeles.

Rose, S., & Walker, R. (2004). A typology of consonant agreement as correspondence. *Language*, 80, 475–531.

Recovering morphology from local phonotactic constraints

Basilio Calderone*, Chiara Celata* and Ivan Herreros**

*Scuola Normale Superiore, Pisa, Italy; {b.calderone, c.celata}@sns.it
**Universitat Pompeu Fabra, Barcelona, Spain

1. Introduction

Implicit phonotactic knowledge emerges from learners' exposure to language-specific distributional information (e.g. Albright & Hayes, 2003; Pierrehumbert & Coleman 1997). 'Wordlikeness' studies (Bailey & Hahn, 2001) distinguish between two different types of knowledge involved in production/judgement of novel words: (a) phonotactic (string level); (b) lexical (word level). Different statistical models are used to shape the two types of knowledge (bi/trigram transition probability for (a) and lexical neighborhood for (b)).

In this work, we propose that the phonotactic level and the word level should not be conceived of as two independent, blind domains, but rather as interconnected systems interacting in a bottom-up fashion. The interplay of phonotactic local constraints and frequency-sensitive word patterns is powerful enough to enable overranked morphological patterns to be recovered from rough data.

2. Experiments. Materials and methods

In inflectional languages, such as Italian, function morphemes are predominantly inserted by suffixation and they tend therefore to occupy the right edge of the word. We hypothesize that, in such languages, similarity effects will emerge more salient at the right edge of a word than in other positions, all other things being equal.

A corpus of 120 trisyllabic bi-morphemic non-words, was created, starting from 20 Italian true morphological endings. For each ending, two sets of 4 pseudo-words were created: one pivot item, and three associated items. In the first set (called Final Positional Option, FPO), the pivot item was created by adding the relevant ending to an arbitrary root morpheme (e.g. ending: *sto#*, pivot item: *ferasto*). Each associated shared the pivot's morphological ending in either final, internal or internal non-adjacent position (e.g. *milusto*, *lustomi*, *sultimo*). The three associated items were composed by the same segments (in the example above, 3 matching and 4 non-matching w.r.t their pivot). The second set of pseudo-words (Non-Final Positional Option, NFPO) was created following the same criteria, but here the selected function morpheme was placed in word initial position (e.g. pivot item *stofera*; associated items *stomilu*, *lustomi* and *sultimo*).

The hypothesis predicts that the similarity patterns of associations within FPO will be significantly different from those occurring within NFPO, overall (Hyp1) and across the three different association conditions (Hyp 2). Evidence comes from a SOM-based computational simulation and a psycholinguistic experiment on native Italian speakers, tested on the same materials. A correlation is expected between the association function given by the simulator and the speakers' similarity rating (Hyp 3).

2.1 Computational evidence

A Self-Organizing Map (SOM) (Kohonen, 1995) is an artificial neural network which uses unsupervised learning to produce low-dimensional representation of the training samples while preserving the topological properties of the input space, that is projecting data points from an n-dimensional input space onto a two-dimensional output space where similar input tokens are mapped onto nearby output units. In this study, a SOM is trained on trigrams extracted from a phonologically encoded spoken Italian corpus. By repeated exposure to the trigram-based training, the map develops memory traces of the most familiar phonotactic patters in an incremental way. In the representation buffer, word forms are re-coded on the basis of acquired phonotactic knowledge.

The activation values (vector modulus) of each morpheme were checked. Morphemes in initial position (e.g. #*sto*) showed ad a significantly higher activation value than the same morphemes in final position (*sto#*), and this clearly reflected the frequency-based different saliency of the two phonotactic contexts.

The similarity values between a pivot and each of its associates were then measured. The similarity between two words *a* and *b* was measured as the cosine between *a* and *b* (with values bounded between 0 and 1). The resulting cosine values were normalized w.r.t. the activation value of each form. The first expectation (Hyp1) was clearly confirmed by our data. Despite the significantly lower value of activation frequency, morphemes in final position showed significantly higher values of similarity association. We interpreted this result as evidence of the fact that, in a language such as Italian, inflectional relations in word final position are strong enough to override salient similarity relations in word initial position and to allow grouping of word units on a morphological ground. On the other hand, Hyp2 did not receive any statistically reliable support, since for both initial and final morphemes, the first association pattern elicited significantly higher similarity values than the two others.

2.2 Psycholinguistic evidence

30 subjects were asked to give a similarity judgment for each pair of pseudo-words, using a 10-point scale. Ratings were then normalized w.r.t the activation value of each form. Similarly to what we found in the SOM-based simulation, final morphemes tended to show higher values of similarity association (Hyp1). No interaction between the association condition and the positional option was found, as in the SOM simulation (Hyp2).

3. Correlation computational/behavioural, and conclusions

The correlation degree between the mean similarity value obtained from speakers' rating and the cosine value obtained from the SOM's similarity association was finally calculated. Consistently with Hyp3, the two variables proved significantly and positively correlated, in particular for some of the morphemes, but altogether as well (r = .5491, $p < .01$). Figure 1 illustrates this result.
Together the SOM simulation and the behavioural evidence showed that phonotactic constraints coupled with frequency information may have significant effects at the word level, where morphological patterns emerge as the by-product of the local representations of chunks of individual elements composing the phonological chain.

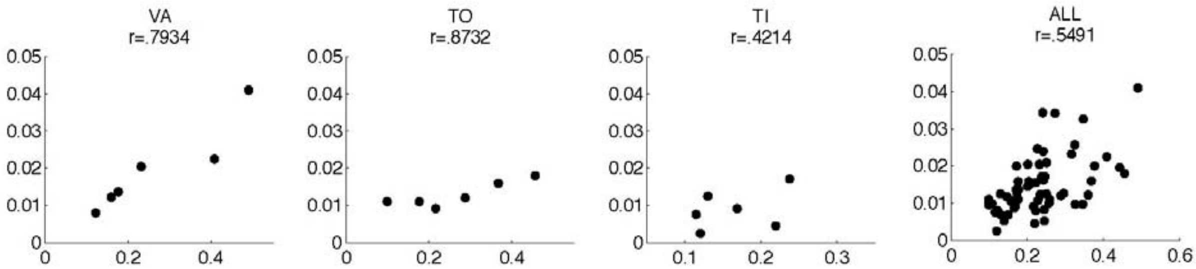


Figure 1: Correlation between SOM's similarity values and subjects similarity ratings across a sample of morphemes, and overall

References

Albright, A., & Hayes, B. (2003). Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition*, 90, 119-161.

Bailey, T.M., & Hahn, U. (2001). Determinants of Wordlikeness: Phonotactics or Lexical Neighborhoods? *Journal of Memory and Language*, 44,568-591.

Coady, J. A., & Aslin, R.N. (2004). Young children's sensitivity to probabilistic phonotactics in the developing lexicon. *Journal of Experimental Child Psychology*, 89,183-213.

Coleman, J. & Pierrehumbert, J. (1997). Stochastic Phonological Grammars and Acceptability. In *3rd Meeting of the ACL Special Interest Group in Computational Phonology*, Proceedings of the Workshop, 12 July 1997. Association for Computational Linguistics: Somerset, 49-56.

Kohonen, T. (2002). *Self-Organizing Maps*. Springer: Berlin.

Lexical Bias in Cross-Dialect Word Recognition in Noise

Cynthia G. Clopper,* Janet B. Pierrehumbert# & Terrin N. Tamati**

*Department of Linguistics, Ohio State University, USA; clopper.1@osu.edu
#Department of Linguistics, Northwestern University, USA; jbp@northwestern.edu
**Department of Linguistics, Ohio State University, USA; tamati.1@osu.edu

Lexical bias is a well-known factor affecting phonological categorization in spoken word recognition tasks. Ganong (1980) observed that listeners shifted their phonetic category boundary for /t/ vs. /d/ depending on the lexical status of the endpoint stimulus materials (e.g., *task* vs. *dask* or *tash* vs. *dash*) to produce more real word responses. The "Ganong effect," therefore, reflects an advantage for real word responses over nonword responses in lexical identification tasks (Fox, 1984; Pitt, 1995). Dialect variation has also been found to affect the perception of phonetic category boundaries. For example, Rakerd and Plichta (2003) found that listeners shifted their phonetic category boundary for /ɑ/ vs. /æ/ depending on the dialect of the preceding carrier phrase. For carrier phrases with fronted /ɑ/s, the boundary was shifted towards /æ/ relative to the boundary for carrier phrases with backed /ɑ/s.

The current study examined the interaction between lexical bias and dialect variation in a spoken word recognition task in noise. The stimulus materials were real English CVC words produced by three female talkers from each of two regional varieties of American English (North and Midland). A set of predicted phonological confusions was established for vowels based on previous descriptions of regional dialect variation in American English (e.g., Labov, Ash, & Boberg, 2006). For example, in the Northern dialect, /ɛ/ is lowered and backed and may be confusable with /æ/. In the Midland dialect, /ow/ is fronted and may be confusable with /ʊ/. To manipulate lexical bias in the current experiment, two sets of stimulus words were constructed. In the first set, the target words were selected so that the predicted phonological confusions resulted in real English words (word competitor condition). For example, all of the selected words containing /ɛ/ had a real minimal pair in English containing /æ/ (e.g., *bet* and *bat*). In the second set, the target words were selected so that the predicted phonological confusions did not result in real English words (nonword competitor condition). For example, none of the selected words containing /ɛ/ had a real minimal pair in English containing /æ/ (e.g., *chess* and **chass*). Each stimulus set included five words for each of 11 vowels in American English (i, ɪ, ej, ɛ, æ, ɑ, ɔ, ʌ, ow, ʊ, u). For the vowels for which no confusions were predicted (i, ej, u), the same words were used in both the word and nonword competitor conditions. The listeners heard each word mixed with speech shaped noise at a signal-to-noise ratio of +2 dB and responded by typing the words that they heard. Nineteen undergraduate linguistics students served as participants in the word competitor condition and 20 undergraduates served as participants in the nonword competitor condition. The residential history of the listeners was mixed, but most were lifetime residents of either the Midland (N=20) or the Northern (N=9) dialect.

The responses were scored for both correct word and correct vowel. Average word and vowel accuracy scores for each talker dialect in each condition are shown in Table 1. At the word level, performance was more accurate for the Northern talkers than the Midland talkers across both lexical conditions. However, at the vowel level, the effect of talker dialect was not significant in either condition, suggesting that while the Northern talkers were more intelligible overall, this intelligibility benefit may have been limited to consonants.

Table 1. Percent correct words and vowels in each condition for each talker dialect.

	Words		Vowels	
Condition	**Midland**	**North**	**Midland**	**North**
Word Competitor	63	67	81	82
Nonword Competitor	71	74	87	88

In addition, both word and vowel recognition performance were more accurate in the nonword competitor condition than the word competitor condition for both talker dialects, consistent with the Ganong effect. When the predicted phonological confusions were real words, listeners could respond with either the target word or the confusable minimal pair. When the predicted phonological confusions were not real words, however, the listeners exhibited a lexical bias and responded with the target word. This interpretation is further supported by an analysis comparing the responses to the words for which no phonological confusions were predicted and which were shared across the two conditions (i, ej, u) to the responses to the words for which phonological confusions were predicted and that differed across the two conditions (ɪ, ɛ, æ, ɑ, ɔ, ʌ, ow, ʊ). At both the word and vowel level, when the no phonological confusions were predicted, performance did not differ between the word and nonword competitor conditions. However, when phonological confusions were predicted, performance was significantly better in the nonword competitor condition than the word competitor condition at both the word and vowel levels, suggesting a lexical bias in the nonword competitor condition.

An examination of the responses to specific vowels further confirmed the Ganong effect and also revealed the role of dialect variation in eliciting this effect in natural speech. In the word competitor condition, /ɪ/ was misidentified as /ɛ/ and /ɛ/ was misidentified as / æ / significantly more often than chance for the Northern talkers, but not for the Midland talkers. This pattern of errors is consistent with the Northern dialect of American English, which includes backing and lowering of both /ɪ/ and /ɛ/. In addition, in the word competitor condition, /ow/ was misidentified as /ʊ/ significantly more often than chance for both the Midland and Northern talkers. This error pattern is consistent with back vowel fronting observed for women in both dialect regions. None of these error patterns were significant in the nonword competitor condition. Thus, predicted phonological confusions were observed in the word competitor condition, when an alternative real English word was available to the listeners as a possible response, but not in the nonword competitor condition, when a real English word was not available as a possible response.

Taken together, these results confirm a lexical bias in spoken word recognition. In addition, whereas previous studies of lexical bias have typically relied on synthetic acoustic continua (e.g., Fox, 1984; Ganong, 1980), the findings in the current study suggest that lexical bias can also be observed for more naturally occurring variability. In particular, vowel variation in American English leads to continuous acoustic-phonetic variability across dialects that is similar to the acoustic continua produced in the laboratory and used for studying category boundaries, and it is exactly these continua of naturally-occurring variation that elicited the effects of lexical bias in the current study.

Acknowledgements

The authors would like to thank John Pate for assistance with data collection.

References

Fox, R. A. (1984). Effect of lexical status on phonetic categorization. *Journal of Experimental Psychology: Human Perception and Performance*, 10, 526–540.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, 6, 110-125.

Labov, W., Ash, S., & Boberg, C. (2006). *Atlas of North American English*. New York: Mouton de Gruyter.

Pitt, M. A. (1995). The locus of the lexical shift in phoneme identification. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 21, 1037–1052.

Rakerd, B., & Plichta, B. (2003). More on perceptions of /A/ fronting. Paper presented at *New Ways of Analyzing Variation 32*, Philadelphia, PA. October 9-12.

The Internal Structure of Nasal-Stop Sequences: Evidence from Austronesian

Abigail C. Cohn* & Anastasia K. Riehl#

*Department of Linguistics, Cornell University, United States; acc4@cornell.edu
#Department of Linguistics, University of Toronto, Canada; anastasia.riehl@utoronto.ca

The phonological and phonetic structure of nasal-stop sequences has elicited much attention. Yet, in fact, less is known about the internal timing of nasal-stop sequences, both unary cases (most commonly prenasalized stops–^{N}D) and clusters (nasal voiced-stop clusters–ND, and nasal voiceless-stop clusters–NT) than often assumed. This calls into question certain phonological assumptions and conclusions. In this paper, we examine the phonetic structure of nasal-stop sequences to address this lacuna. (*Nasal-stop sequence,* or *NC sequence,* is used here to refer to both unary and cluster cases.)

In a cross-linguistic study of the realization of nasals and nasalization in English, French, and Sundanese, Cohn (1990) observed a systematic asymmetry in the relative timing of the nasal and oral portions in nasal-voiceless stop vs. nasal-voiced stop cases: For the NT cases, the nasal and oral components each take up about half of the total duration, as expected; while in the ND cases, the sequence is nasal for all but a very brief period. While others have since noted a similar asymmetry, no full account has been offered. The goal of this paper is to better understand this asymmetry and in so doing also consider the status of so-called "postploded nasals" (N^{D}) (Blust, 1997). We present data from six Austronesian languages to investigate these issues. The Austronesian language family is known for its rich array of nasal-stop sequences, including cases of both prenasalized stops and what have been described as postploded nasals.

We present some background on the question of prenasalized stops vs. nasal-stop clusters, reviewing relevant results from Riehl (2008), before turning to the nature of the asymmetry between the voiced and voiceless NCs and the status of postploded nasals. We will see that in each of these three cases the central issue is timing: first, total duration is the crucial dimension that allows us to differentiate unary segments vs. clusters; second, relative timing of the nasal and oral components is critical to characterizing the difference between nasal voiced- and nasal voiceless-stop sequences; third, micro-timing (structure of the transition and nature of the oral component) is at issue in assessing whether postploded nasals are distinct from other sorts of nasal-stop sequences.

We present acoustic and nasal airflow data for six Austronesian languages, which between them are described to exemplify four nasal-stop sequence-types: ^{N}D, ND, N^{D}, and NT. Erromangan and Tamambo (Oceanic languages of Vanuatu) are described as having prenasalized stops. Acehnese and Sundanese (West Austronesian languages of Indonesia) are described as having postploded nasals. Manado Malay and Pamona (also West Austronesian languages of Indonesia) are described as having nasal voiced-stop clusters. In addition, in each of these languages except for Tamambo, the nasal voiced-stop sequences contrast with nasal voiceless-stop clusters. The data consist of target words with nasal-stop sequences, in minimal sets with corresponding plain nasals and stops, recorded in appropriate frame sentences; results are presented for multiple repetitions for 1-6 speakers per language. Measurements are presented for the absolute and relative duration of the components: nasal portion, oral portion, and burst.

Overall timing

In terms of overall timing, contra recent claims in the literature, Riehl (2008) finds strong support for the conclusion that there is a systematic difference in overall timing between unary cases (^{N}D) and clusters (ND) whereby ND is substantially longer than ^{N}D. She concludes that the best way to assess the difference is by looking at a ratio of the duration of a plain nasal to the duration of an NC sequence. In addition, based upon a cross-linguistic survey of NCs, Riehl proposes a classification of occurring and non-occurring patterns. She finds two clear gaps: first, no language has unary ^{N}T; second, no language exhibits a contrast within morphemes between ^{N}D and ND, unless the language independently has a singleton-geminate contrast. These two gaps account for the fact that languages have, at most, a two-way contrast: ND vs. NT or ^{N}D vs. NT.

Relative timing

In terms of duration of the component parts, our data consistently reveal the same asymmetry observed by Cohn (1990). The duration of the oral and nasal portions is roughly as expected in nasal voiceless-stop clusters for all five languages. However, for all the nasal voiced-stop cases, the nasal component takes up all but the final portion of the combined total duration. Notably, this difference holds whether the ND sequence is unary or cluster. Finally, there are some fine-grained timing differences in the relative duration of the oral transition and release between the nasal voiced-stops (discussed below).

Not only the differences in relative timing between ND and NT in each language, but also the similarities of ^{N}D, N^{D}, ND across languages require an account. We conclude that the explanations lie in the phonetics, not the phonology, since there is strong phonetic consistency across languages independent of phonological structure. First, in the ND cases, only the nasality changes, but in the NT cases, both nasality and voicing change. To accomplish this, NT requires an abrupt change from nasal to oral and a clear oral closure, given the precise timing needed to change both voicing and nasality. In the ND cases, we argue that the combined effects of the lack of perceptual saliency of nasality during closure, the strong propensity for nasalization to persevere, and the difficulty of changing from spontaneous (nasal) voicing to obstruent voicing during the closure account for the internal structure of these sequence-types.

Micro-timing

Fine-grained differences in the relative duration of the oral transition and release in the nasal voiced-stop cases were observed. This leads us to consider whether the differences between nasal-stop clusters and prenasalized stops and so called postploded nasals lie in these phonetic details. To address this, we further investigated the details of the oral closure and accompanying burst in these six cases. We find that the languages fall along a continuum from having a clear oral closure and burst, to only a burst, to no burst. However, these differences do not relate directly to the putative phonological status of the sequences. For example, Manado Malay, described to have ND clusters, shows the least oral structure. We also find inter-speaker variation, which in some cases is highly systematic.

This leads to two questions: 1) how the contrast between N and ND is realized and 2) whether ^{N}D and N^{D} are distinct. First, even in cases where no clear burst is present, there are two key differences between the ND sequence-types and plain nasals. These are all languages that exhibit robust carryover vowel nasalization. Perceptually, the most salient cue to whether the consonant is a plain nasal or a nasal voiced-stop sequence is whether the following vowel is nasalized or oral. Second, there is a systematic duration difference, whereby postploded nasals in Acehnese and Sundanese are longer than the plain nasals (confirming Durie's 1985 observation for Acehnese), as is also the case for ND cluster cases. This directly reflects the fact that these sounds are actually ND clusters rather than unary segments. Thus the difference between ^{N}D and reported cases of N^{D} lie in the overall timing difference between unary and cluster cases. At the same time, the inconsistent correlation between micro-timing properties and ND clusters vs. N^{D} calls into question whether these are actually distinct cases. The only systematic difference observed among the ND sequence-types is the overall timing difference between the unary and cluster cases.

In conclusion, we find that a fuller understanding of the phonetic timing and structure of NC sequences in turn informs our understanding of phonological patterns and systems of contrast.

References

Blust, R. (1997). Nasals and nasalization in Borneo. *Oceanic Linguistics,* 36, 149-179.

Cohn, A. (1990). *Phonetic and Phonological Rules of Nasalization.* Ph.D. Dissertation, UCLA. [*UCLA Working Papers in Phonetics 76*]

Durie, M. (1985). *A Grammar of Acehnese on the Basis of a Dialect of North Aceh.* Dordrecht, Holland: Foris Publications.

Riehl, A. (2008). *The Phonology and Phonetics of Nasal-Obstruent Sequences*. Ph.D. Dissertation, Cornell University.

Frequency and repetition effects outweigh phonetic detail in prominence perception

Jennifer Cole*, Yoonsook Mo* & Mark Hasegawa-Johnson#

*Department of Linguistics; #Department of Electrical Engineering
{jscole; ymo; jhasegaw}@uiuc.edu
University of Illinois at Urbana-Champaign,USA

This paper reports on a study of prominence perception focused on the question of whether words that are perceived as prominent are distinguished by acoustic cues to prosody (signal-driven perception), and/or whether the listener may be guided by non-phonetic information, including information related to the discourse context of the utterance, and lexical knowledge of word-specific phonetic detail derived from the listener's prior experience (expectation-driven perception). We are especially interested to find out if the discourse and lexical factors influence prominence perception even in the absence of strong prosodic cues in the acoustic signal. The analysis is based on data from our large-scale study of naïve prosody perception (Mo, Cole, & Lee, 2008), in which untrained subjects are asked to listen to excerpts from spontaneous, conversational speech, and to mark a printed transcript of the speech in real-time, by underlining words that are perceived as prominent. Prominence was defined as a highlighting function used by the speaker for the listener's benefit, to make some words stand out relative to others. No illustrative examples of prominence were given, and transcribers saw no graphical display of the speech waveform, spectrogram or pitch track. (A second task involved transcribing the location of prosodic "chunk" boundaries, but those results are not discussed here.) The speech materials consist of short excerpts (20-30 sec.) representing 20 speakers from the Buckeye Corpus of spontaneous American English as spoken in Columbus, Ohio (Pitt, Dilley, Johnson, Kiesling, Raymond, Hume, & Fosler-Lussier, 2007). Each excerpt is independently transcribed by multiple transcribers (between 15-22 transcribers per excerpt, 72 transcribers total), and though there is substantial variation in perceived prominence across speakers and across transcribers, the results of inter-transcriber agreement show highly significant agreement rates based on Fleiss' statistic. Transcriptions are pooled and each word in every excerpt is assigned a P(rominence)-score, which is a number from 0 to 1 representing the proportion of transcribers (out of 15-22) who perceived that word as prominent.

Our prior work (Mo 2008) establishes significant correlations between P-scores and the acoustic measures of duration and intensity: words that are perceived as prominent exhibit stressed syllables with significantly greater duration, and for some vowels, greater intensity and bandpass filtered intensity (spectral emphasis). This finding is consistent with ***signal-driven prominence perception***, as these acoustic features may directly encode prosodic prominence and therefore serve as primary cues for prosodic context. The present paper introduces new evidence from multiple linear regression analyses showing significant correlations between P-scores and the non-signal-based predictor variables of **lexical frequency** and **discourse repetition**. Lexical frequency is measured here as the log frequency of each word based on its occurrence in the Switchboard corpus of American English conversational speech. The repetition factor encodes prior mention of the word in the discourse excerpt (first- *vs*. second- *vs*. subsequent-mention).

Results of linear regression show a significant inverse correlation between log-frequency and P-scores (N= 2024, R^2=.26, p<0.001) when function words are included, with similar findings in an analysis that excludes function words. Further, considering now only content words that are repeated within the same discourse excerpt (N=176), there is a smaller but still significant inverse correlation between the repetition factor and P-score (R^2=.059, p<0.002). These findings show that the rate of perceived prominence is lower for frequent words and for words that are repeated in the discourse. And while this finding introduces the notion that lexical and discourse factors may play a role in prominence perception, it does not immediately challenge the hypothesis of signal-driven prominence perception. After all, prior studies have shown that high-frequency and repeated words undergo reduction similar to the reduction observed in non-prominent prosodic contexts. Reduction due to frequency or repetition also results in lower duration and intensity values. Thus, if prominence perception is guided primarily or exclusively by properties of the acoustic signal, as predicted by the signal-driven prominence perception hypothesis, we would still expect high-frequency words and repeated words to be correlated with lower P-scores.

Additional regression analyses were conducted in order to test the hypothesis of ***expectation-based prominence perception***, i.e., that lexical frequency and discourse repetition have a direct effect on the listener, independent of the information in the acoustic signal. Lexical frequency and discourse repetition were tested as predictors for two acoustic measures often associated with prominence: duration and intensity (overall and bandpass filtered in three frequency bands) of the primary stressed vowel. The results show significant effects of lexical frequency on overall RMS intensity (R^2=.024, p=0.041) and RMS intensity in each of three frequency bands (0-500 Hz, R^2=.034, p=0.014; 500-2000 Hz, R^2=.028, p=0.028, 2000-4000 Hz, R^2=.045, p=0.005), but only one marginally significant effect of repetition in discourse, and that was on stressed vowel duration (R^2=.024, p=0 .044).

To summarize, the non-signal-based factors of lexical frequency and discourse repetition are significantly correlated with perceived prominence in our data: high-frequency and repeated words are less likely to be judged as prominent than low-frequency or non-repeated words. We note that the combined influence of frequency and repetition on prominence perception is not very strong, accounting for less than 32% of the variance in P-scores, which suggests that there may be additional factors influencing prominence perception that have not yet been considered. Our second finding, as predicted from previous studies, is that lexical frequency and discourse repetition are also correlated with acoustic measures of intensity and duration: high-frequency words have lower intensity, while repeated words exhibit somewhat shorter duration. We note, however, that frequency and repetition are more strongly correlated with prominence perception than they are with the intensity and duration values associated with prominence. This interesting finding supports the hypothesis of expectation-based prominence perception. Listeners are influenced in their judgment of prominence by the frequency of a word and its prior mention in the discourse, beyond the degree expected based on the evidence in the acoustic signal. Taken together, the results from this study suggest a model of prosodic prominence that exists in the interaction between speaker and hearer. The listener perceives the prosodic cues produced by the speaker, but may also be 'biased' in their perception of prominence based on their prior experience of the word, its phonetic encoding, frequency, and discourse context. We discuss an interpretation of this perceptual bias as the result of speeded processing: a high-frequency or repeated word has a higher resting activation level, and may thus be processed more quickly by the listener. A word may be judged as 'prominent' only if it is processed more deliberately (using greater perceptual resources), with attention to its phonetic detail. In this way, prominence may be viewed as a phenomenon of speech processing, rather than an attribute of the lexical item.

Acknowledgements

This research is supported by NSF grants IIS 07-03624 and IIS 04-14117. Statements in this paper reflect the opinions and conclusions of the authors, and are not necessarily endorsed by the NSF. Thanks to Zachary Hulstrom, Eun-Kyung Lee, Soondo Baek, and the UIUC Prosody & ASR Research Group.

References

Mo, Y. (2008). Duration and intensity as perceptual cues to naïve listeners' prominence and boundary perception. Paper presented at *Speech Prosody 2008,* Campinas, Brazil.

Mo, T., Cole, J., & Lee, E. (2008). Naïve listeners' prominence and boundary perception. Paper presented at *Speech Prosody 2008,* Campinas, Brazil.

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).

Nasalization of /æ/ and sound change in Australian English

Felicity Cox* & Sallyanne Palethorpe#

* Department of Linguistics, Macquarie University, Australia; f.cox@ling.mq.edu.au
#Macquarie Centre for Cognitive Science, Macquarie University, Australia; sassi@maccs.mq.edu.au

In Australian English (AusE), a phonetic lowering of the short front series of vowels /æ, e, ɪ/ (TRAP, DRESS, KIT) is currently in progress. This change began with /æ/ (Cox, 1999) which was previously located in the vicinity of cardinal 3 [ɛ] but now occupies the extreme open position of the inverted triangular vowel space (Cox & Palethorpe, 2007). We would expect this change to encompass the various contextually based realizations through a shift in the exemplar distribution (Pierrehumbert, 2003). According to exemplar theory, words and contexts that have a high probability of occurrence (token frequency) skew the distribution in the direction of the change (Bybee, 2004). Therefore, the lowest /æ/ variants should occur in high frequency words. High frequency phonotactic sequences (type frequency) might also affect distribution skewness (Hay et al., 2003). One highly probable phonotactic environment for /æ/ is the pre-nasal context (Vitevitch & Luce, 2004) which is also the context most likely to inhibit lowering of this vowel. Numerous studies have shown that nasalized /æ/ is phonetically raised relative to its oral counterpart (Beddor, 1993).

In the current AusE /æ/ category, the opposing forces of nasalized raising combined with the probabilistic lowering of the change may create a pull in different directions and could result in an elongated distribution representing an increase in the gradient of realizations for this vowel from the phonotactically probable nasal through to the most lexically frequent oral tokens.

Based on predictions from exemplar theory, speakers should exhibit phonetically open distributions of /æ/ across contexts as a result of the sound change and high frequency words containing non-nasal consonants should occur at the lower edge of the distribution. We would also expect words containing vowels in nasal contexts to have phonetically closer realizations in line with research on nasalization. We would expect the distribution to be contained within the general /æ/ space to maintain contrast with /e/, however gradient expansion resulting from antagonism between type frequency and the direction of the change is predicted.

In order to examine these predictions, we conducted an acoustic analysis of the vowels /æ/ and /e/ in the speech of 15 lower middle class AusE speaking female university students from Sydney's northern districts. All were speakers of General AusE and under 25 years of age. Recordings were made in a sound treated studio in the Centre for Language Sciences at Macquarie University. Speakers read from a computer screen, three times in random order, the 12 monophthongs of AusE in the standard hVd frame and also the vowels /æ/ and /e/ in the monosyllabic and disyllabic contexts /CVn, CVd, CVnd, CVdə, CVnə, CVndə/ where C was a range of consonants with differing place, manner and voicing characteristics including /p, b, t, d, k, h, s/. A subset of the data containing the CVn and CVd words has been selected for this analysis. The formant frequencies of F1 and F2 as well as vowel durations were extracted using standard procedures described in Cox (1999). The values for the three repetitions of each word were averaged for each speaker to overcome the problem of artificially inflating the dataset. Mixed model analysis was used to examine the effect of word type and nasality on the formant and duration values.

Results show that word type did not affect formant values indicating that lexical characteristics were not responsible for the distributions. All speakers produced open oral /æ/ and phonetically more raised. nasalized /æ/. However, only one third of the speakers had a gradient oral to nasalized /æ/ within the /æ/ category space as predicted. The majority produced a categorical distinction in phonetic height between the oral and nasalized vowel such that the nasalized /æ/ patterned with /e/. Cluster analysis confirmed the presence of two groups of speakers: those who substantially raised nasalized /æ/ (categorical group) and those who did not (gradient group). Figure 1 illustrates these effects for syllables containing onset /b/. Results also show that duration is used to maintain the contrast between the nasal allophones of /e/ and /æ/.

This recent sound change in AusE /æ/ appears regular as no evidence of lexical diffusion was found in the analysis. Phonetic context rather than lexical effects were shown to have the major influence. These results support Labov's (2006) findings for North America that such vowel shift changes are not influenced by lexical frequency.

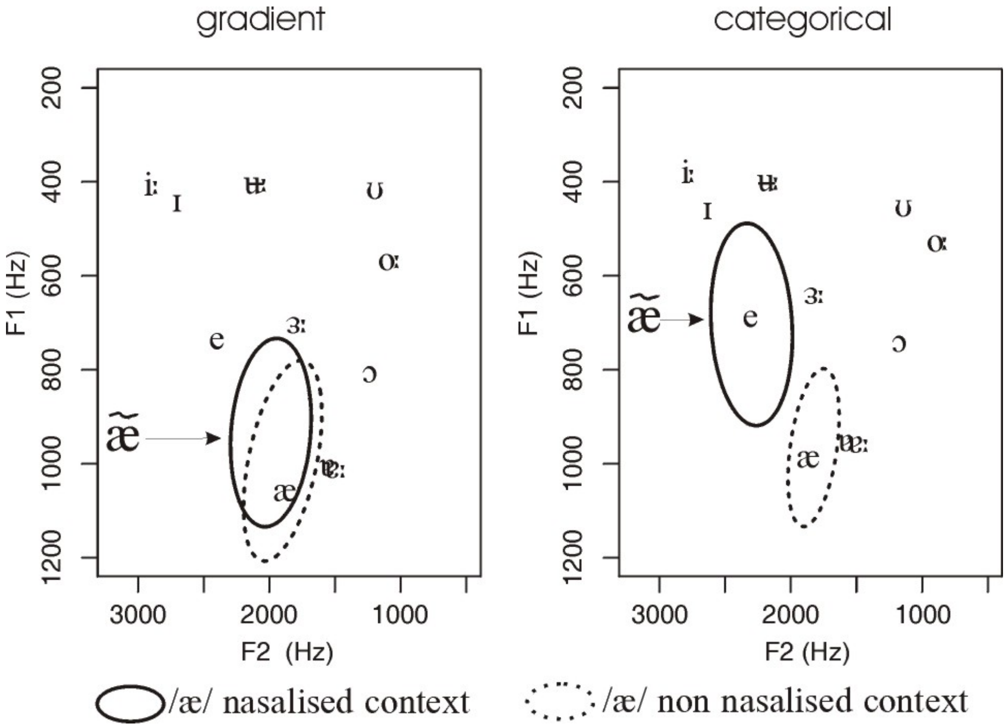


Figure 1: Oral and nasalized /æ/ ellipses for speakers who produce gradient distributions (left panel) and those who produce categorical distinctions (right panel).

Acknowledgements

We would like to thank Chris Callaghan, John Lonergan and Kimiko Tsukada for assistance with data collection and labelling.

References

Beddor, P. (1993). The perception of nasal vowels. In M. K. Huffman & R.A. Krakow (Eds.), *Phonetics and Phonology: Nasals, Nasalisation and the Velum* (pp. 171-196). San Diego: Academic Press.

Bybee, J. (2004). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change,* 14, 261-290.

Cox, F. (1999). Vowel change in Australian English, *Phonetica,* 56, 1-27.

Cox, F., & Palethorpe, S. (2007). Illustration of the IPA: Australian English. *Journal of the International Phonetic Association,* 37, 341-350.

Hay, J. Pierrehumbert, J., & Beckman, M. (2003). Phonetic interpretation. In J. Local, J. Ogden & R. Temple (Eds.) *Laboratory Phonology 6* (pp. 58-74). Cambridge: Cambridge University Press.

Labov, W. (2006). A sociolinguistic perspective on sociophonetic research, *Journal of Phonetics*, 34, 500-515.

Pierrehumbert, J. (2003). Probabilistic-phonology: discrimination and robustness. In R. Bod, J. Hay & S. Jannedy (Eds.), *Probabilistic Linguistics* (pp. 177-228). Cambridge: MIT Press.

Vitevitch, M.S., & Luce, P.A. (2004). A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, and Computers,* 36, 481-487.

Phonetic variation and phonological phrasing: does the Accentual Phrase exist in Italian?

Mariapaola D'Imperio & Caterina Petrone

Université d'Aix-Marseille I & Laboratoire Parole et Langage, CNRS, France
mariapaola.dimperio@lpl-aix.fr, caterina.petrone@lpl-aix.fr

Within the Autosegmental-Metrical (AM) Theory of intonation (Pierrehumbert, 1980, inter alia), tunes are represented as a sequence of one or more pitch accents followed by a phrase accent and a boundary tone. Phrase accents and boundary tones generally occur at the end of phrasal constituents, such as the intermediate, the intonation or the accentual phrase. Nuclear accents are positionally defined as the last and more prominent accents in the intermediate phrase, immediately preceding a phrase accent. In Standard Italian, the nuclear accent has been defined as "the rightmost fully-fledged pitch accent in the focussed constituent" (Grice et al., 2005, pg. 380). Moreover, the number and the definition of the phrasing levels in this language are still quite controversial. In fact, whereas both the intonation and the intermediate phrase are well attested in many Italian varieties (Grice et al., 2005), there is yet no tonal evidence for the existence of phrasing constituents smaller than the intermediate phrase. However, within a different and less recent approach (Nespor & Vogel, 1986), it has been proposed that a prosodic constituent immediately below the intermediate phrase might be tonally marked in Italian, i.e. the phonological phrase *(φ)*, which includes an independent lexical head plus all its complements on its non-recursive side (e.g. *la mamma*, "the mom", *il benevolo manovale* "the benevolent worker"). Moreover, evidence for a tonal constituent smaller than the intermediate phrase, the Accentual Phrase (*AP*), whose domain roughly correspond to that of φ, has been found for example in Korean (Jun, 1993) and French (Jun & Fougeron, 2002), or even in stress languages such as Arabic (Hellmuth, in press).

In this work, acoustic data are reported in which we report on evidence for the tonal marking of a prosodic domain which might be similar to that of the *AP*. Specifically, this constituent appears to be marked at its right edge by a tone which is differently specified for yes-no questions and statements, i.e. a H edge tone for questions vs. a falling L tone for statements. We already know that in Neapolitan the nuclear LH rise is later in yes/no questions (L*+H) than in narrow focus statements (L+H*) (D'Imperio, 2000). Moreover, in long focus constituents the F_0 contour following the prenuclear rise falls rapidly in statements while it stays high in questions (D'Imperio, 2002, 2003). From informal observations, it appears that the F_0 contour within the region spanning from the prenuclear to the nuclear accent is differently realized in the two modalities. Specifically, in statements the F_0 rapidly falls from the H prenuclear peak to the region around the end of the host *(φ),* whereas in questions the F_0 fall after the prenuclear peak is shallower, so that the F_0 contour in the immediate postaccentual region assumes a concave shape.

To verify whether such tonal differences are systematic, a pilot study was conducted in which a set of sentences was read by two Neapolitan speakers. Each sentence was composed of three constituents: an utterance-initial *φ* bearing a (L)H* prenuclear accent (*la mamma*, "the mom"); an unaccented verbal phrase (*vuole vedere*, "wants to see") and an utterance-final *φ* (es. *la Bina*, "(the) Bina", proper name) bearing a rising nuclear accent. The sentences were read either as yes/no questions or as (late) narrow focus statements. The predicted difference in shape and slope of the region immediately following the prenuclear rise (the first LH in the left and right panel of Fig. 1) was actually found, confirming our observations. Specifically, in statements after the prenuclear rise a low turning point can be located around the end of *la mamma* (Fig. 1, right panel), followed by a low plateau continuing until the beginning of the nuclear rise. In questions, on the other hand, the postaccentual region is characterized by a concave downward parabola, whose inflection point is also located around the end of *φ*. After this region, the slope becomes steeper in order to attend the low values for the L* of the L*+H nuclear accent (Fig. 1, left panel).

Results from tonal alignment and scaling suggest that this difference is not due to phonological/phonetic variation in the neighboring tones. Since interpolation between tonal targets is assumed to be linear within the AM approach, we therefore asked whether the difference could be accounted for by the presence of an edge tone after the postaccentual rise, which would intervene between the H prenuclear peak and the L nuclear tone. Results from the combined use of linear (linear piecewise regression) and non linear (logistic) modelling techniques suggest that a tonal target is present at the end of *la mamma*, and that this tone is differently specified for statements (L) and questions (H), thus mirroring some regularities already found in Neapolitan long focus constituents. We also argue that such a difference in tonal specification would help to enhance the phonological contrast between questions and statements, for which no morphosyntactic differences are exploited and whose main cue is the different alignment of the nuclear accent. Note also that in both intonation modalities, this phrasal tone was not accompanied by the percept of a break, as found at the end of the intermediate phrase (*ip*) or intonational phrase (*IP*).

In order to assess whether a constituent smaller than the *ip* can be tonally marked in Italian, we ran a second study in which we manipulated the intonation modality (questions vs. statements) and the absolute position of the target word (*benevolo,* "benevolent") within the intonation phrase and the intermediate phrase (final vs non final). Preliminary results from two out of four speakers confirm our expectations: when the target word is in non-final position within the *ip*, a phrasal tone occurred at its right edge, which is differently specified depending on intonation modality. Moreover, in both questions and statements, the length of the target word was progressively longer in non-final *ip*, final *ip* and final *IP* position, with different degrees of perceptual break accompanying the end of the three constituents. Our findings confirm the existence of a third level of phrasing in Neapolitan smaller than the *ip* and the IP, at the same time raising some problems for approach to intonational meaning which do not take into account the interrelation of tones within a tune.

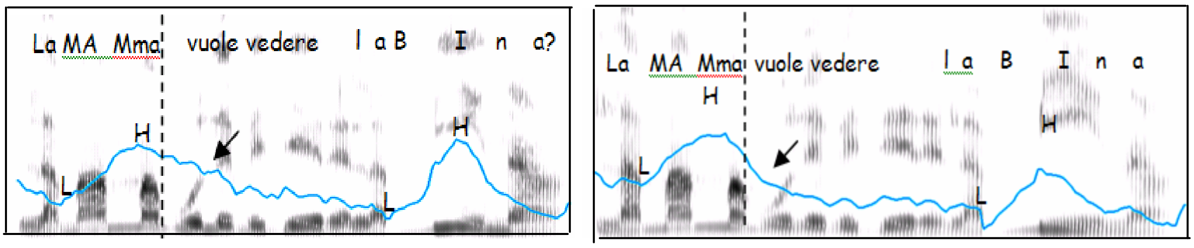


Figure 1: F_0 contour of a yes/no question (left) and narrow focus statement (right) utterance for the sentence *La mamma vuole vedere la Bina* "The mother wants to see (the) Bina". The dotted line marks the end of the first φ. The region following the prenuclear rise is indicated by the black arrow.

References

D'Imperio, M. (2000). *The Role of Perception in Defining Tonal Targets and their Alignment.* PhD Thesis, The Ohio State University.

D'Imperio, M. (2002). Italian intonation: An overview and some questions, *Probus* 14(1), 37-69

D'Imperio, M. (2003). Tonal structure and pitch targets in Italian focus constituents, *CatJL*, 2, 55-65.

Grice, M., D'Imperio, M., Savino, M., & Avesani, C. (2005). Towards a strategy for labelling varieties of Italian. In S.-A. Jun (Ed.) *Prosodic Typology and Transcription: A Unified Approach* (pp. 55-83). Oxford: OUP.

Hellmuth, S. (in press) Prosodic weight and phonological phrasing in Cairene Arabic. *Proceedings of Annual Meeting of Chicago Linguistic Society* 40.

Jun, S.-A. (1993). *The Phonetics and Phonology of Korean Prosody*. Ph.D. dissertation, OSU.

Jun, S.-A., & Fougeron, C., (2002). The Realizations of the Accentual Phrase in French Intonation. *Probus*, 14, 147-172.

Nespor, M., & Vogel, I. (1986). *Prosodic Phonology*. Foris: Dordrecht.

Pierrehumbert, J. B. (1980). The Phonology and Phonetics of English intonation. PhD Thesis, MIT.

The Gradient Phonotactics of English CVC Syllables

Olga Dmitrieva & Arto Anttila

Department of Linguistics, Stanford University, USA; dmitro@stanford.edu

This study examines the factors affecting the gradient well-formedness of English CVC syllables. We will focus on two major effects: place dissimilation between the two consonants and the prominence alignment between stress, vowel height, and consonant place. First, recent research has shown that the Obligatory Contour Principle (OCP) emerges gradiently in phonotactics. In particular, Berkley (1994) demonstrated a gradient OCP-place effect in the English lexicon: monomorphemic monosyllabic words where the onset and the coda share place of articulation are statistically underrepresented. We replicate and extend Berkley's result using CELEX2 (Baayen, Piepenbrock, & Gulikers, 1995) and CMU (Weide, 1998) dictionaries, showing that the gradient OCP-place effect emerges not only in monomorphemic monosyllabic words, but in all CVC syllables of English. Second, we point out a gradient prominence alignment effect among stress, vowel height, and consonant place, which requires that the most prominent element in each domain combines with the most prominent elements in other domains.

In our study, we code each CVC syllable in the CELEX and CMU databases for (1) the place feature of the onset and the coda (dorsal, labial, and coronal), (2) vowel height/sonority (low and mid vs. high and reduced), and (3) stress of the syllable (stressed and unstressed). The degree of the representation of each syllable type in the lexicon is quantified in terms of the Observed/Expected frequency ratio (Frisch, Pierrehumbert, & Broe, 2004). Expected frequency was calculated by multiplying the probabilities of each term, based on the assumption that they are independent. Resulting O/E values show that the CVC syllables with the homorganic consonants in onsets and codas are underrepresented in the dataset compared to the syllables with the heterorganic onsets and codas. We also observe that low and mid vowels are underrepresented in combinations with coronal consonants and in unstressed syllables, while high and reduced vowels are underrepresented in combinations with labial and dorsal consonants and in stressed syllables.

To evaluate the size of the effect, we apply a multiple regression analysis (Coetzee & Pater, 2008). The prohibition against low/mid vowels in unstressed syllables and high/reduced vowels in stressed syllables emerges as the most robust factor ($p < 0.001$). The OCP effect emerges as the second most predictive factor ($p < 0.001$). Finally, the prohibition against marked consonants (dorsal, labial) in syllables that were unstressed or had low-sonority vowels (high, reduced) also reached significance ($p < 0.01$). This effect is more pronounced for onsets than for codas. Interestingly, the symmetric counterpart of the last factor, that is, the prohibition against unmarked consonants (coronal) in syllables that were stressed or had high-sonority vowels (low, mid) vowels was not significant. These results suggest that vowels in English CVC syllables undergo gradient neutralization in non-prominent positions and augmentation in prominent positions, while consonants undergo neutralization alone. Thus, stressed syllables achieve both acoustic salience (through more sonorous vowels) and informational efficiency (through an extended consonantal inventory). Unstressed syllables, on the other hand, reduce the vowel and consonant inventory to the least prominent segments.

The study suggests a number of further questions. The most important one is the possible effect of the inflectional and derivational morphology on the statistical patterns in the data. In particular, phonological and prosodic characteristics of inflectional endings and derivational affixes could be responsible for boosting certain quantitative patterns, such as overrepresentation of coronal consonants in the unstressed syllables. A preliminary investigation using the CELEX corpus involved excluding all final syllables that are likely to carry most of the morphological suffixes. Excluding morphological suffixes did not appear to cause significant changes in the observed patterns. This suggests that that the presence of morphologically complex items in the dataset only serves to magnify the prominence alignment effects, but is not the source of these effects. A more detailed investigation of the morphological factors is left for the future work.

The study has one particularly interesting theoretical consequence. The relationship between vowel sonority and syllable prominence is well attested, and phonetically grounded. Stress is often attracted to acoustically more prominent (longer, louder) low and mid vowels. Vowels can undergo augmentation in stressed syllables and reduction in unstressed syllables. The connection between consonant place and prominence, on the other hand, is controversial. De Lacy (2001) argues that constraints relating prominence and place of articulation, e.g. *STRESS/LABIAL, cannot exist since no phonological alternations of this kind are attested. Our study challenges this assumption by showing that there is a connection between the place feature of the consonants and syllable prominence. We find a statistically significant preference for coronal onsets and codas over dorsal and labial onsets and codas in unstressed CVC syllables.

There is also some evidence that the connection between consonant place and prominence may be phonetically motivated. Compared to dorsals and labials, coronals require less articulatory effort and time. At least for coronal stops, in some cases this leads to less overlap with neighboring sounds and lower perceptibility. In contrast, dorsal and labial consonants involve larger and slower moving articulators, which results in more coarticulation in the adjacent sounds and enhanced perceptual salience (Jun, 2004).

We propose an optimality-theoretic account for the data, which models the gradient well-formedness of syllables in terms of their relative grammatical complexity measured in terms of ranking information (Anttila, 2008). Grammatical complexity is a direct consequence of the relative markedness of the syllable: the more ranking information is required for the syllable to surface faithfully, the more grammatically complex it is. The model predicts that the grammatically most complex syllables should be least well-formed and least frequent, and the less grammatically complex the syllable is, the more well-formed and more frequent it should be. We show that the model correctly predicts a number of quantitative asymmetries in the data. The grammar benefits significantly from the results of the statistical analysis: the constraints were chosen so as to correspond to the factors that reached significance in the regression analysis, e.g. *X/i 'Avoid a combination of stressed syllable and high vowel' and the rankings among the constraints were introduced based on their factors' coefficients in the regression analysis. As a result, the advantages of the statistical techniques combined with the devices of the formal grammar make the proposed approach adequate from both a descriptive and explanatory standpoints.

References

Anttila, A. (2008). Gradient phonotactics and the Complexity Hypothesis. To appear in *Natural Language and Linguistic Theory*.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX Lexical Database* (Release 2). Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania [Distributor].

Berkley, D. (1994). The OCP and gradient data. *Studies in the Linguistic Sciences,* 24, 59–72.

Coetzee, A., & Pater, J. (2008). Weighted constraints and gradient restrictions on place co-occurrence in Muna and Arabic. To appear in *Natural Language and Linguistic Theory*.

de Lacy, P. (2001). Markedness in prominent positions. In O. Matushansky, A. Costa, J. Martin-Gonzalez, L. Nathan, & A. Szczegielniak (Eds.), *HUMIT 2000: MITWPL 40* (pp. 53-66). Cambridge, MA: MIT Working Papers in Linguistics.

Frisch, S. A., Pierrehumbert, J. B., & Broe, M. B. (2004). Similarity avoidance and the OCP. *Natural Language and Linguistic Theory*, 22(1), 179-228.

Jun, J. (2004) Place assimilation. In B. Hayes, R. Kirchner, & D. Steriade (Eds.), *Phonetically-based Phonology* (pp. 35-57). Cambridge: Cambridge University Press.

Weide, R. (1998). *The CMU pronunciation dictionary* (Release 0.6). Carnegie Mellon University. Available online at http://www.speech.cs.cmu.edu/cgi-bin/cmudict.

Perceptual evaluation of sociophonetic variability: how do listeners learn?

Gerard Docherty*, Christian Langstrof* & Paul Foulkes#

*School of Education Communication & Language Sciences, Newcastle University, UK;
g.j.docherty@ncl.ac.uk; c.langstrof@ncl.ac.uk
#Department of Language and Linguistic Science, University of York, UK; pf11@york.ac.uk

There is now fairly widespread acceptance that the representation of the phonological shape of words in memory takes a hybrid form, consisting in part of memory traces of tokens of those words experienced by listeners, and that these traces encode gradient phonetic detail alongside a range of contextual factors intrinsic to particular exemplars (e.g. details of speaker, place, context, etc.). One of the attractive features of a lexical representation which is phonetically rich is that it provides a basis for accounting for the learning, use and evaluation of social-indexical variability as an integral part of speech performance and speech perception (Foulkes & Docherty, 2006; Pierrehumbert, 2006), but there has been very little attempt to date to explore the mechanics of how the mapping of phonetic correlates to social category/ies actually happens.

In the present study, we set out to test the hypothesis that exposure to phonetic variability which is distributed such that particular phonetic forms are associated entirely or predominantly with particular social groups should lead listeners, over time, to form associations in memory between those forms and the relevant social category. Previous work by Foulkes et al. (forthcoming) tested this hypothesis using natural data and focusing on the fact that listeners in Newcastle (UK) are exposed to gender-differentiated variants of /t/ (for example, pre-aspirated /t/ is much more frequent in female speech performance than in male speech). In a word-list task, listeners were asked to judge the sex of androgynous (child) speakers, and our analysis focused on whether responses to words with plain vs preaspirated vs glottal variants of /t/ were skewed by listeners' experience with these variants. While the results were to a degree compatible with the hypothesis, there were a number of difficulties with the particular method adopted and with extending that methodology to a wider range of variables.

In this study, we have adopted a different approach in order to investigate the process through which participants learn social-indexical associations. The method involves training participants on novel patterns of sociophonetic variability and, in a subsequent test phase, assessing the extent to which this training leads participants to generate new associations between specific realisational variants and particular social categories, as the exemplar account of lexical representation suggests that they should. Our method enables us to address questions such as whether certain social-indexical properties of speech are easier to become attuned to than others, how much exposure is needed for an individual to link a particular pattern of variation to a novel social category, how categorical does a phonetic variant/social category association have to be in order for it to be learned, and how consistent is cross-individual performance in this sort of learning?

This paper reports the results from 27 participants distributed over three conditions. In each condition, participants were first of all exposed to a training set of stimuli which incorporated novel associations between particular phonetic realisations and social categories. The stimuli conveying the novel sociophonetic associations were merged randomly with an equal number of control words with a similar phonological structure but encapsulating no particular systematic variability in relation to the novel social categories which the speaker is being exposed to. The stimuli were produced by four speakers cross-balanced across all of the criterial and control training stimuli. Stimuli were presented to listeners over headphones using the DMDX display software. In the subsequent test phase of the experiment, listeners heard the same material and were asked to respond to each stimulus as rapidly as they could, indicating which social category they believed the speaker producing the stimulus belongs to. The labels used for the social categories were kept as neutral as possible, with the stimuli in the training phase being identified as being produced by a speaker from either *tribe1* or *tribe2*.

We investigated three conditions:

condition 1 (6 participants): disyllabic words with intervocalic /t/; tribe 1 stimuli always had plain [t]; tribe 2 always had [?]; disyllabic control words with intervocalic stops other than /t/.

condition 2 (15 participants): as in condition 1, except that the weighting of the plain vs. glottal variant was 80:20 in tribe 1 stimuli; tribe 2 had a 20:80 weighting.

condition 3 (6 participants): monosyllabic words with the FLEECE vowel; tribe 1 stimuli always had monophthongal [i]; tribe 2 had a slightly diphthongised variant; monosyllabic control words with a range of monophthongs.

Our findings to date suggest novel sociophonetic associations can be acquired on the basis of exposure to material which embeds that association even if the association in the training takes the form of a (strong) tendency rather than a categorical association (conditions 1 and 2). On the other hand, listeners appear to be unable to learn novel associations relating to fine-grained (but systematic) variation in vowel variants (condition 3). We discuss these results and their implications and point to further work, now under way, in which we systematically vary the various parameters which we can control (e.g. nature of the variants and degree of listener familiarity with the same, skewness of variant distribution across social categories, number of social categories, amount of training, time lapse between training and testing, etc.) in order to develop a deeper understanding of the factors which underpin the learning of novel social-indexical properties of speech.

References

Foulkes, P. & Docherty, G.J. (2006). The social life of phonetics and phonology. *Journal of Phonetics*, 34(4), 409-438.

Foulkes, P., Docherty, G.J., Khattab, G., & Yaeger-Dror, M. (forthcoming). Sound judgements: perception of indexical features in chidren's speech. To appear in Dennis Preston & Nancy Niedzielski (Eds.) *A Reader in Sociophonetics*. Berlin: Mouton de Gruyter.

Pierrehumbert, J.B. (2006). The next toolkit. *Journal of Phonetics*, 34(4), 516-530.

Sensitivity to Grammatical and Sociophonetic Variability in Perception

Katie Drager

Department of Linguistics, University of Canterbury, New Zealand; katiedrager@gmail.com

Sociolinguistic ethnographies provide insights into how individuals manipulate both linguistic and non-linguistic variants to construct their identities, and they allow examination of social categories beyond those traditionally employed (Eckert, 1989). Recent research using experimental methodologies has shown that social and phonetic information are linked in the mind; social information attributed to a speaker can be influenced by altering which phonetic variants perceivers are exposed to, even when differences between the variants are extremely subtle (Campbell-Kibler, 2006). In addition to socially-conditioned variation, there is evidence that phonetic realizations in production vary depending on the grammatical function of a word (Plug, 2005; Hay & Bresnan, 2006), yet it is not known whether listeners can actively exploit this phonetic variation in speech perception. This paper reports on two perception experiments conducted to determine whether perceivers' sensitivity to fine phonetic detail can assist in extraction of both grammatical and social meaning from the signal.

During a year-long ethnographic study conducted at an all girls' high school in New Zealand, I observed a binary division among the girls depending on where they ate lunch: in the common room (CR) or not (NCR). Acoustic phonetic analysis was conducted on the girls' tokens of the word *like*, which has different grammatical functions including a lexical verb (I quite LIKE you[1]), a discourse particle (I was just LIKE singing[2]), and a quotative (she was LIKE whoa)[3]. The analysis reveals variation in whether or not the /k/ is realized. A binomial mixed effects model, with speaker as random effect, reveals a significant interaction between /k/-realization and where the speaker eats lunch; a token where the /k/ is present is less likely to be quotative *like* if produced by a CR girl, and more likely to be quotative *like* if produced by a NCR girl.

To determine the extent to which the girls are sensitive to this subtle relationship between social, grammatical and phonetic variation, several speech perception experiments were conducted. 28 participants (14 CR, 14 NCR) completed three related experiments in which short clips from spontaneous speech produced by the participants and their classmates were used as stimuli. In the first two experiments, the participants were asked to identify the grammatical function of an auditory token of *like*, and in the third experiment, they were asked whether an auditory token of *like* was produced by a girl who eats lunch in the CR or not. This paper reports on results from experiments 1 and 3.

In experiment 1, the question of interest was whether participants could use phonetic detail to distinguish between grammatical functions of *like*. For each question, participants were played two short clips, each containing a token of *like* (e.g. *I was like*), one of which was the quotative and the other of which was the discourse particle. They were asked to indicate which grammatical context on the answer sheet they believed each token had been taken from. A binary mixed effects model was fit to the data from experiment 1 modeling accuracy on matching the auditory token of *like* with its context. Question number and participant were included as random effects in the model. The model includes intonation contours in the stimulus and social information of the speaker. It also includes a significant interaction between /k/ presence in the stimuli and whether a participant produces the /k/ in non-quotative functions of *like*; girls who are more likely to realize the /k/ in non-quotative *like* relative to each other are more accurate when the stimulus quotative has the /k/ present than when it does not. The girls who are most likely to realize the /k/ in non-quotative *like* are even more likely to realize the /k/ in quotative *like*. Likewise, the girls who are least likely to realize the /k/ in non-quotative *like* are even less likely to realize the /k/ in quotative *like*. Given these

[1] Rose, CR
[2] Rochelle, CR
[3] Isabelle, NCR

production trends and the make-up of the stimuli, this interaction is interpreted as a reflection of a link between an individual's production and their perception.

For the third experiment, the question of interest was whether participants could use fine phonetic detail to identify social information, namely, the lunch habits of the speaker. In task 1, participants were told they would hear a token of lexical verb *like*, and in task 2, participants were told they would hear a token of quotative *like*. For task 3, participants were exposed to both lexical verb *like* and quotative *like.* Upon hearing a token, they were asked to indicate whether they believed the speaker ate lunch in the CR or not. A binary mixed effects model was fit to the data, modeling whether a voice was accurately identified as eating lunch in the CR or not. Question number and participant were included as random effects in the model. Included as fixed effects in the model are whether the voice was correctly identified and whether the stimulus voice belongs to a girl who eats lunch in the CR; participants are more accurate if they correctly recognize the voice, and there is an overall bias toward responding that the voice belongs to a CR girl. Also included is whether the stimulus token of lexical verb *like* has the /k/ present; participants are more accurate if a question includes a token of lexical verb *like*, especially if that token does not have the /k/ present. In general, NCR girls are more likely to drop the /k/ in lexical verb *like* than CR girls, and both stimulus tokens of lexical verb *like* where the /k/ is dropped were produced by NCR girls. Overall, participants were biased toward indicating that the speaker ate lunch in the CR, but they were more likely to correctly identify a girl as NCR if the stimuli included a token of lexical verb *like* where the /k/ is dropped. In other words, participants were more accurate on the task when the stimuli best reflect lemma-conditioned sociophonetic patterns of speech observed at the school.

Altogether, the results from these experiments provide evidence that individuals are sensitive to grammatical and sociophonetic variation during perception; very subtle social and grammatical variability must be indexed to fine-grained phonetic detail in the mind. These results are consistent with an exemplar model of speech production and perception where acoustically-rich exemplars are indexed to the social information of a speaker (Johnson, 1997). The results display the benefits of employing an ethnographic approach alongside experimental methodologies, and they demonstrate the value of using natural, spontaneous speech which is familiar to the participants as experimental stimuli.

Acknowledgements

I would like to thank Jen Hay and Alex D'Arcy. I would also like to thank the girls at Selwyn High.

References

Campbell-Kibler, K. (2006). *Listener Perceptions of Sociolinguistic Variables: The case of (ING).* PhD thesis. Stanford.

Eckert, P. (1989). *Jocks and Burnouts: Social Identity in the High School*. New York: Teachers College Press

Hay, J., & Bresnan, J. (2006). Spoken Syntax: The phonetics of giving a hand in New Zealand English. *The Linguistic Review*, 23 (3), 321-349.

Johnson, K. (1997). Speech Perception without speaker normalization: An exemplar model. In K. Johnson, J. W. Mullennix (Eds.), *Talker Variability in Speech Processing* (pp. 145-65). San Diego: Academic Press

Plug, L. (2005). Phonetic reduction and categorisation in exemplar-based representation: Observations on a Dutch discourse marker. In S. Blaho, L. Vicente & E. Schoorlemmer (Eds.), *Proceedings of ConSOLE XIII*.

Experimental approaches to lexical representations of tone in Swedish: What is stored and how does it affect lexical word access?

Verena Felder*, Elisabet Joensson-Steiner*, Carsten Eulitz* & Aditi Lahiri#

*Department of Linguistics, University of Konstanz, Germany; verena.felder@uni-konstanz.de
#Department of Linguistics, University of Oxford, England; aditi.lahiri@ling-phil.ox.ac.uk

Half the languages in the world use tonal cues in addition to segments for differentiating between words. In the investigation of word identification, it is of interest to consider the relative contribution of suprasegmental elements in comparison to that of segments. Studies on tone in Chinese (Cutler & Chen, 1997; Ye & Connine, 1999; Schirmer, Tang, Penney, Gunter & Chen, 2005) and Japanese (Cutler & Otake, 1999; Sekiguchi & Nakajima, 1999) have shown that tone plays an important role in accessing words in the mental lexicon. An incorrect tone can narrow down the cohort of activated words – sometimes as effectively as do mismatching segments.

Germanic languages usually do not have lexical tone. Curiously, some Scandinavian languages, like Swedish and Norwegian, are classified as tonal languages. They have a binary tone contrast, referred to as Accent 1 (henceforth ACC1) and Accent 2 (ACC2). The tonal contrast is only observable in words that minimally contain a trochee – i.e. a stressed syllable followed by an unstressed syllable. Acoustically ACC1 begins with a rise on the first syllable, whereas ACC2 has a fall. Additionally, ACC2 has a second peak in the post-stressed syllable. Since ACC2 can be realized only on trochees, monosyllabic words are ACC1 by default, even if they are pronounced as disyllabic words due to postlexical processes (e.g. */humr/* > *hummer*). Theories on the lexical representation of the binary tonal contrast usually assume privative specification of ACC2, because of its' higher complexity. Contrary to this common view, we assume that ACC2 is assigned by default to any underlyingly disyllabic word not specified for ACC1.

In contrast to the Asian tonal languages investigated so far, in Swedish there are very few monomorphemic minimal pairs in the language distinguishing words solely by means of tone. In principle, a lack of tone should not hamper word identification. This calls into question the role tone plays in lexical access in Swedish. As a first step, we designed a forced choice study to investigate Swedish listeners' sensitivity to tonal contrasts and their ability to identify the tone with the acoustic information of the first syllable only. Stimulus material consisted of 60 pairs of disyllabic Swedish nouns. The first syllables in each pair had the same segmental structure, but the words differed in accent (e.g. *<u>ham</u>bo*$_1$ – *<u>ham</u>pa*$_2$). The items belonged to one of four stimulus groups: (a) ACC1-S: disyllabic words assumed to be specified for ACC1 (e.g. *hambo*$_1$), (b) ACC1-U: underlyingly monosyllabic words that surface as trochees but carry ACC1 by default (e.g. *hummer*$_1$), (c) ACC2-CS: the disyllabic ACC2 words with the same first syllables as the ACC1-S words (-CS meaning "corresponding to ACC1-S") (e.g. *hampa*$_2$) (d) ACC2-CU: the disyllabic ACC2 words segmentally corresponding to ACC1-U (e.g. *humla*$_2$). No differences were expected between ACC2-CS and ACC2-CU, both are underspecified for tone and receive default ACC2. They were split into –CS and -CU to control for potential effects of segments.

In the forced choice task, participants were presented auditorily with the first syllable of an item (e.g. *hum-*) and then saw both words of the pair on the screen (e.g. *HUMMER HUMLA*). Their task was to decide which word the auditory fragment was taken from. Auditory stimulus material was gained by cutting the first syllable of each word. In a gating study with removed pitch information, it was made sure that decisions could not be taken on grounds of segmental/coarticulatory information. Consequently, the task could only be achieved by extracting the tonal information. Response accuracy revealed that correct tone could be identified by hearing only the first syllable. The tone of ACC1-U and ACC1-S words was recognized slightly better than that of the ACC2-CU and ACC2-CS words. This is probably caused by the fact that the second peak of ACC2 was missing in the signal. Interestingly enough, response speed showed a different picture. ACC1-S words were classified significantly faster, while there was no difference in reaction time between the latter three groups. Frequency, cohort size and word length could be ruled out as possible causes. We hypothesized that the processing of surface tonal contours was governed by their underlying lexical specification, in the sense that words with specified ACC1 in the mental lexicon are identified faster than words not specified for tone. This contradicts the assumption of specified ACC2 but instead speaks for the lexical specification of disyllabic ACC1 words (ACC1-S).

Studies on tone so far have generally considered the relative importance of segments vs. tone, but none included hypotheses on lexical specifications of tone in a given language. In an electroencephalographic (EEG) cross-modal fragment priming study we (a) attempted to clarify the relative contribution of Swedish tone as compared to segments and (b) explored effects of lexical specification on brain responses. The same stimuli were used. Again, the first syllable of a word was presented auditorily as the prime fragment. It was followed by a visually displayed target word. Participants had to decide whether the target was a word or not. The prime fragment was either identical to the target word in segments and tone (e.g. ham_1-HAMBO_1), differed in tone but not in segments (e.g. ham_2-HAMBO_1), differed in segments but not in tone (e.g. bul_1-HAMBO_1), or was completely different, sharing neither tone nor segments (e.g. bul_2-HAMBO_1). The first question was, whether a fragment can preactivate a target if it is only identical in segments, but differs in tone. If tone is a strong cue in the language, it should be considered in lexical access and should be able to rule out alternative candidates. Results show that a mismatch in segments did not preactivate the target, no matter whether tone was same or different. That is, as expected, tone by itself is not sufficient to activate a tonal cohort ignoring the segments. When the segments were identical for prime and target, ACC1 as well as ACC2 words showed a difference in amplitude between same and different tone. This confirms that tone has an impact in lexical access in Swedish.

Concerning lexical specification, in response to the target words, ERPs differed in amplitude between ACC1-S and the rest. This is the same pattern as we had in the first experiment. Also, brain responses elicited by ACC1-U did not differ from those of ACC2, although the former words are underlyingly monosyllabic. This was true irrespective of the preceding prime. Between priming conditions, the factor of lexical specification revealed no significant results. Therefore, no conclusions on the impact of tonal lexical specification on priming can be drawn in this experiment.

Acknowledgements

This research was made possible by funding from the Deutsche Forschungsgemeinschaft (German Research Foundation, SFB471).

References

Cutler, A., & Chen, H.-C. (1997). Lexical Tone in Cantonese spoken-word processing. *Perception & Psychophysics*, 59, 165-179.

Cutler, A., & Otake, T. (1999). Pitch accent in spoken-word recognition in Japanese. *Journal of the Acoustical Society of America*, 105, 1877-1888.

Schirmer, A., Tang, S.-L., Penney, T. B., Gunter, T. C., & Chen,H.-C. (2005). Brain responses to segmentally and tonally induced semantic violations in Cantonese. *Journal of Cognitive Neuroscience*, 17, 1-12.

Sekiguchi, Nakajima (1999), The use of lexical prosody for lexical access of the Japanese language. *Journal of Psycholinguistic Research,* 28, 439-454.

Ye, Y., & Connine, C. M. (1999). Processing spoken Chinese: The role of tone information. *Language and Cognitive Processes*, 14, 609-630.

Front/Back Asymmetries in Height Harmony

Sara Finley* & William Badecker*

*Department of Cognitive Science, Johns Hopkins University, Baltimore MD;
finley@cogsci.jhu.edu; badecker@jhu.edu

This paper argues for substantive biases in phonological generalization using data from an artificial grammar learning experiment involving vowel height harmony. Vowel height harmony is a process whereby vowels are either raised or lowered in order to agree with the height of an adjacent vowel. While the process of vowel height harmony appears in a wide number of Bantu languages (Hyman, 1999) and a number of Romance varieties (Penny, 1970), vowel height harmony is much more restrictive than other types of vowel harmony (Linebaugh, 2007). In Bantu, the majority of height harmony languages show a front-back asymmetry in which front vowels are more likely to undergo harmony than back vowels (Hyman 1999). In this paper, we provide evidence for the substantive nature of this asymmetry. If the learner is equipped with both a bias for front vowel height harmony and a dispreference for back vowel height harmony, it is possible that learners of a vowel height harmony language will apply the height harmony rule to front vowels more often than to back vowels, even if the learners are exposed only to back vowel alternations during training. Our results provide support for these biases– adult English learners of a height harmony language were more likely to apply a height harmony rule to suffixes containing front vowels than suffixes containing back vowels, even when participants were trained only on back vowel suffixes.

The front/back asymmetry in height harmony arises from the phonetic markedness of alternations involving /u/ to [o]. This cross-linguistic avoidance is explained by the fact that [o] is phonetically more marked than [u]; the higher the round vowel, the less marked it is, as cues for rounding decrease as the height of the vowel decreases. This general phonetic markedness is increased when front, unround vowels trigger height harmony because the perceptibility of [o] is particularly weak after front, unround vowels (Kaun, 2004). These factors can lead to a strong dispreference for back vowels undergoing height harmony. If learners of a height harmony language share this general dispreference, then we expect it to be reflected in their performance– greater adherence for front vowel alternations than back vowel alternations.

Participants in our experiment (n = 60; 20 in each condition) were exposed to a stem-controlled height harmony system in which harmonic stems (e.g., [pidu, gobe]) induced alternations in either a front vowel suffix ([-gi]/[-ge]) (Front Suffix Training) or a back vowel suffix ([-gu]/[-go]) (Back Suffix Training). Training consisted of exposure to 24 stems, each immediately followed by their concatenated form (e.g., [pidu, pidu-gi]), repeated 5 times each. All stems contained the vowels [i, e, u, o] and the consonants [p, t, k, b, d, g, m, n]. All consonant and vowel combinations occurred with equal frequency. Following training, participants were tested via a forced choice task: one item contained a harmonic suffix ([pidu-gi]) the other item contained a disharmonic suffix (*[pidu-ge]). Test items were divided into three types (12 items for each type): Old Stems/Old Suffix, which contained items identical to those heard at training, New Stems/Old Suffix, which contained novel stem items, but the identical suffix vowel, and Old Stems/New Suffix Vowel, which contained the same stems as heard in training, but with a reversal in backness from training (e.g., if trained on [-gi]/[-ge], were tested with [-gu]/[-go]).

Results showed a strong preference for harmonic responses for front vowel suffixes. In the Front Vowel Training condition, there was a significant effect of training for Old items and New Stem items, but not for New Suffix Vowel test items ($t(38) < 1$), indicating learning for front vowel suffixes, but a lack of generalization to back vowel suffixes. Interestingly, participants showed an effect of learning for front vowel suffixes even when participants were trained on back vowels. In the Back Vowel Training condition, participants showed no effect of training for Old Stems or New Stems, but showed a significant effect of training for New Suffix Vowel items ($t(38) = 3.58$; $p < 0.01$), indicating an effect of learning for novel suffix vowels. These results suggest a strong preference for front vowel height harmony over back vowel height harmony. Note that Control participants showed no preference for front vowel suffixes over back vowel suffixes, suggesting that participants learned a harmony rule that was more robust for front vowels. Our results support the hypothesis that learners are biased towards front vowel height harmony.

To ensure that our results were not due to acoustic differences in F1 in the test stimuli, we compared the F1 values for front and back vowels. High vowels had significantly lower F1 values for both front and back vowels ($F(1, 32) = 147.28$; $p < 0.001$) but there was no interaction ($F < 1$), indicating that the degree of F1 difference for high and mid vowels is the same for both front and back vowels. Thus, the preference for front vowel harmony alternations is not an artifact of our stimuli.

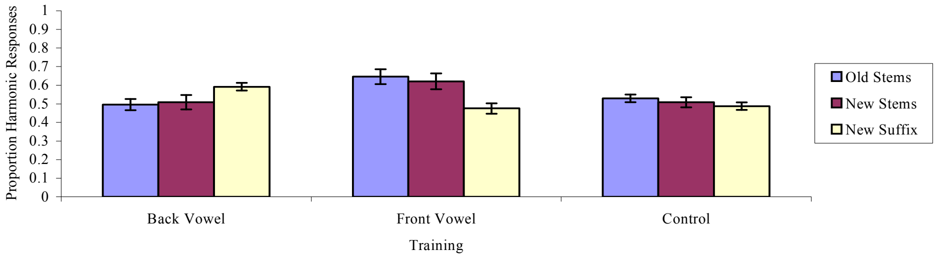


Figure 1: Proportion of Height-Harmonic Responses

The results of our experiment pose a unique problem for both exemplar (e.g., Nosofsky, 1986; Pierrehumbert, 2000) and abstract, categorical theories of processing (e.g., Optimality Theory (Prince & Smolensky, 1993/2004)). Exemplar models predict that the highest level of performance should be for items specifically trained on, but our results show minimal evidence of learning for trained-on back vowel items, but strong evidence of learning for novel front vowel test items. An abstract model with a bias for front vowel height harmony (e.g., front vowel height harmony outranking back vowel height harmony in an Optimality Theoretic model) can provide an account of the differential treatment for front and back vowels in height harmony. However, such an account cannot explain the gradience of learning data (i.e., non-categorical performance across participants and test items), and why height harmony learning data is less robust than back vowel harmony learning data (Linebaugh & Cole, 2005; Pycha, Nowak, Shin, & Shosted, 2003). We propose a model of learning that takes into account both the strengths of the exemplar models and the OT models. The model incorporates substantive biases for vowel harmony typologies that regulate the learnability of different harmony structures, explaining the difference in robustness of back harmony compared to height harmony.

Acknowledgements We would like to thank Paul Smolensky, Colin Wilson, Ariel Goldberg, Rebecca Morley, and members of the JHU Psycholinguistics and IGERT laboratories.

References

Hyman, L. H. (1999). The historical interpretation of vowel harmony in Bantu. In J.-M. Hombert & L.M. Hyman (Eds.) *Bantu HistoricalLinguistics: Theoretical and empirical perspectives* (pp. 235-395). Amsterdam: John Benjamins

Kaun, A. (2004). The typology of rounding harmony. In B. Hayes & D. Steriade (Eds.), *Phonetics in Phonology* (pp. 87-116). Cambridge: Cambridge University Press.

Linebaugh, G. (2007). *Phonetic Grounding and Phonology: Vowel backness harmony and vowel height harmony.* University of Illinois at Urbana-Champaign.

Linebaugh, G., & Cole, J. (2005). *Does vowel harmony facilitate speech production?* Paper presented at the McWOP, University of Michigan.

Nosofsky, R. (1986). Attention, similarity and the identification-categorization relationship. *Journal of Experimental Psychology: General,* 115, 39-57.

Penny, R. J. (1970). Mass-nouns and metaphony in the dialects of north-western Spain. *Archivum Linguisticum,* 1, 21-30.

Pierrehumbert, J. (2000). Exemplar dynamics: Word frequency, lenition and constrast. In J. L. Bybee & P. Hopper (Eds.), *Frequency effects and emergent grammar* (pp. 137-157). Amsterdam: John Benjamins.

Prince, A., & Smolensky, P. (1993/2004). *Optimality theory: Constraint interaction in generative grammar*. Cambridge: Blackwell.

Pycha, A., Nowak, P., Shin, E., & Shosted, R. (2003). Phonological rule-learning and its implications for a theory of vowel harmony. *WCCFL, 22*, 101-113.

Exploring social-indexical knowledge: a long past but a short history

Paul Foulkes

Department of Language and Linguistic Science, University of York, UK; pf11@york.ac.uk

Language functions in a social context. All languages vary systematically as a function of their social roles, and in response to their users. Such facts were acknowledged by Panini around 600 BC, and also by the neogrammarians and structuralists, to which modern linguistics of course owes a great deal. For theoretical reasons, however, the dominant models of linguistic theory in the twentieth century steered away from social matters, relegating issues of variation and social indexicality from serious consideration.

One of the defining themes of linguistic research over the last few years has been an emerging consensus that a cognitively-realistic, integrated theory of phonological knowledge, speech production, and speech perception must include more than an account of those properties pertaining to lexical contrast. It is clear that what speaker-listeners know about language involves not only abstract symbolic representations of linguistic structures, but also an extensive repository of indexical information. The latter includes knowledge about how to decode and interpret indexical information when it is encountered in listening, and how to encode it within speech production to signal aspects of personal identity and to achieve pragmatic goals in the course of speaking. Moreover, there is also a growing body of evidence that indexical knowledge and linguistic knowledge interact with each other. A central plank in the case for this consensus has been the widespread acceptance of some role for episodic or exemplar models of memory, and usage-based statistical learning. Support for this position has come from a remarkably wide spectrum of sources, including speech perception and psycholinguistics (e.g. Pisoni, 1997), phonetics (Pierrehumbert, 2003), sociolinguistics (Foulkes & Docherty, 2006), phonology (Bybee, 2001), syntax (Jackendoff, 2007), L1 acquisition (Vihman & Croft, 2007), and L2 acquisition (Cutler & Weber, 2007). It now seems uncontroversial to conclude that social information is retained in memory alongside linguistic knowledge. Questions remain, however, over what sorts of social information are learned and stored, where and how they are stored in relation to linguistic information, and how social information affects linguistic processing.

In considering these issues, I have three main aims in this presentation. First, I offer a review of what we know about social-indexical information at the level of phonetics/phonology, drawing on evidence from a range of disciplines. I suggest that social-indexical information is more complex and more extensive than is typically acknowledged in laboratory phonology. In turn this points to the necessity of sharing elements of methodology and theorising across disciplines.

Secondly, I explore evidence for the development of indexical knowledge. Rather little research has explored how knowledge of indexical variation develops in acquisition. We therefore have a relatively poor understanding of what is learned, when and how it is learned, and of what relationships may obtain between indexical and linguistic knowledge. I review the evidence for indexical variation in the input which children receive, how it is manifested in children's speech production, and how children's perceptual capacities attune to indexical information. I then discuss the evidence from the perspective of exemplar models. One of the attractions of exemplar models is their capacity to predict learning of both linguistic and social structures through the same mechanism. The tenets of exemplar models also make a number of predictions about indexical learning: some types of indexical properties are likely to be transmitted more readily than others, as a function, for example, of the frequency and phonetic transparency with which they are manifested. The predictions appear to be borne out relatively well by the available data. Indexical learning appears to begin even *in utero*, with recognition of the maternal voice apparent in neonates. Learning of familiar voices and phonetically transparent indexical categories (e.g. talker sex) appears earlier than more arbitrary and less frequently manifested categories (e.g. regional dialects).

Inevitably much remains to be explored. Therefore I end by raising a number of implications and challenges for exemplar models. These include the following:

1. Indexical learning appears to precede the development of language (defined in the usual sense). Early indexical learning may thus provide a foundation for the learning of abstract linguistic structures. By virtue

of experiencing the maternal voice through pre-natal experience, the child enters the world with an elementary framework for categorisation, able to distinguish the normal maternal voice from other sounds.

2. Indexical and linguistic knowledge are likely to be intertwined in initial learning (Docherty et al., 2006). Whether they are ever teased apart into two discrete cognitive systems remains a moot point. However, it is clear that indexical information is not totally discarded. With this in mind, a key task in acquisition is to learn to focus on abstract information and indexical detail as appropriate to the context of language use.

3. Different children face different tasks in acquisition. Assuming a bias to maternal input, replete with information indexing gender, boys may face the specific task of diverging from the maternal model.

4. The exemplar view of speech production remains underdeveloped, particularly with respect to weighting and saliency. Given the great range of indexical factors inherent to speech, weighting is likely to be a complex and important process. It may well begin very early. Furthermore, we should not neglect the role of the child himself as an agent in shaping his linguistic destiny. Production biases emerge in infants at around 9 months (Vihman & Croft, 2007), as babbling preferences help to map adult input onto production templates. The child may also focus on his own output as a source of input, reinforcing the effects of the templates.

5. There is abundant evidence that early representations are surface-based, and thus likely to encode indexical detail. Early representations may therefore be some way removed from the citation forms often assumed in developmental studies. Citation forms may not crystallise until orthography is mastered.

6. In laboratory phonology we should not limit our conception of indexicality, and subsequently our research designs, to a simple set of broad categories. What speakers and listeners learn, know and process is much richer than can be revealed by simple comparisons such as males against females. Extending experimental investigation further into 'the field' is likely to be a fruitful policy. There is also much to be learned from a deeper trawl into the literature in adjacent fields such as acquisition, anthropology, dialectology, sociolinguistics, bilingualism, and conversation analysis, for insights into the social structures and indexical categories that have relevance for particular individuals in particular circumstances.

While much remains to be done in understanding indexical knowledge, this is an interesting period in which to be working in phonology. Some have identified it as a new era, a point of paradigm shift (e.g. Hawkins, 2004). Exemplar theory, for all its remaining deficits, provides the foundation to construct a cognitively-realistic theory of phonological knowledge in which indexical knowledge is not marginalised but central. Exploring indexical knowledge may have a short history, but the future appears long and profitable.

References

Bybee, J. L. (2001). *Phonology and Language Use*. Cambridge: Cambridge University Press.

Cutler, A. & Weber, A. (2007). Listening experience and phonetic-to-lexical mapping in L2. In J. Trouvain & W.J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 34-48). Saarbrücken, Germany.

Docherty, G. J., Foulkes, P., Tillotson, J. & Watt, D. J. L. (2006). On the scope of phonological learning: issues arising from socially structured variation. In L. Goldstein, D. H. Whalen, & C. T. Best (Eds.), *Laboratory Phonology 8* (pp. 393-421). Berlin: Mouton de Gruyter.

Foulkes, P., & Docherty, G. J. (2006). The social life of phonetics and phonology. *Journal of Phonetics*, 34, 403-438.

Hawkins, S. (2004). Puzzles and patterns in 50 years of research on speech perception. In J. Slifka, S. Manuel, & M. Matthies (Eds.) *From Sound to Sense: 50+ Years of Discoveries in Speech Communication,* Cambridge, MA: MIT.

Jackendoff, R. (2007). Linguistics in cognitive science: the state of the art. *The Linguistic Review,* 24, 347-401.

Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech,* 46, 115-154.

Pisoni, D. B. (1997). Some thoughts on 'normalization' in speech perception. In K. Johnson & J. W. Mullennix (Eds.), *Talker Variability in Speech Processing* (pp. 9-32). San Diego: Academic Press.

Vihman, M. M., & Croft, W. (2007). Phonological development: toward a 'radical' templatic phonology. *Linguistics*, 45, 683-725.

Metalinguistic judgments of phonotactics by bilinguals

Stefan A. Frisch, María R. Bréa-Spahn & Carla I. Orellana

Department of Communication Sciences and Disorders, University of South Florida, frisch@cas.usf.edu

Introduction

It is well-established that speakers' metalinguistic processing directly reflect the statistical properties of the lexicon (e.g. Treiman et al., 2000; Hay, Pierrehumbert, & Beckman, 2004). Frisch, Large, & Pisoni (2000) found that speakers' judgments of well-formedness was similar whether they were given a 7-point Likert scale to rate 'wordlikeness' or a binary 'acceptable/unacceptable' decision, so there may be no absolute cutoff for grammatical vs. ungrammatical novel phonological forms. Frisch (in prep) found that a measure of vocabulary familiarity predicted individual differences in well-formedness judgments in English speakers, such that participants with relatively larger vocabularies were more likely to accept low probability nonwords as well-formed. This paper demonstrates that a speaker's judgments of well-formedness for novel forms is based on generalizations over lexical knowledge by showing individual differences in performance in metalinguistic processing of novel words by bilinguals.

Data

An experiment was conducted involving 30 fluent bilingual Spanish-English speakers using two tasks: 1) Well-formedness judgments for auditorily presented novel nonwords (Frisch, Large, & Pisoni, 2000), and 2) Familiarity judgments for orthographically presented real words (Nusbaum, Pisoni, & Davis, 1984). There were 120 multisyllabic nonword stimuli for each language, generated by random concatenation of onset and rime units using the probabilistic grammar model of Coleman & Pierrehumbert (1997). The probability distributions are created from a phonotactic analysis of words in a dictionary (the Webster's dictionary of Nusbaum et al 1984 for English and a modified version of the LDC Callhome Spanish lexicon for Latin American Spanish, which excluded morphologically complex words as much as possible to create a lexicon similar to what is typically used for studies of English phonotactics). In the English word familiarity experiment, there were 50 word stimuli from the low to mid familiarity words in Nusbaum et al. (1984). The Spanish familiarity task used 100 randomly selected words from the low to mid frequency range in the LDC Callhome lexicon that were screened for specific dialect usage. It was not possible to make a Spanish familiarity task fully comparable to the English task as there is no database of word familiarity in Spanish that we are aware of.

Findings

A moderate significant correlation between log probability and mean word-likeness rating across participants was found in both English ($r = .44$) and Spanish ($r = .42$) nonwords. Within subjects, individual correlations between log probability and rating were generally positive and significantly greater than zero on average for both languages on a t-test. These findings replicate previous results with English speakers, and extend those results to bilingual speakers using English or Spanish.

Vocabulary effects within language

Individual correlations between word-likeness rating and participants' vocabulary familiarity scores were correlated with phonotactic probability across the nonwords. A significant correlation was found for bilinguals in English ($r = -.16$), replicating previous work with English monolinguals (Frisch, in prep). A negative correlation means that there was more of a correlation between word-likeness rating and vocabulary familiarity for low probability nonwords compared to high probability nonwords. The correlation was not found for bilinguals in Spanish, which may reflect the fact that the Spanish vocabulary test is less well-developed. Post-hoc evaluation of the vocabulary test for Spanish found that when the vocabulary familiarity score is based on the 20 items that are most diverse across participants, a correlation in the predicted direction emerges, though it is still non-significant ($r = -.05$).

Effects across languages

Cross-language phonotactic influences were examined by creating blended grammars that mixed the probabilities of constituents across the two languages to various degrees. Where the two languages had similar categories (e.g. the /i/ vowel), the probabilities of the constituents in each language were blended by taking a weighted average of the probabilities in the two languages. In cases where one language was missing the appropriate category (e.g. the /ɪ/ vowel), the probability for one language was zero, and the effect of blending languages was to reduce the probability of constituents that occurred in one language but not the other, relative to constituents that occurred in both. Language blends were created for each individual participant in order to best fit their performance in the well-formedness task for each language separately. Overall, probability predictiveness was highest within each language separately.

There were 9 of 30 participants where the fit to Spanish nonwords improved if some effect of English probabilities was included. Overall, this effect was negatively correlated with Spanish vocabulary (so those with higher Spanish vocabulary were less likely to benefit from mixing in English probabilities). There was almost no effect of English vocabulary, but all of the participants with high Spanish vocabulary and low English vocabulary did not benefit from mixing English with Spanish to predict Spanish nonwords.

There were 10 of 30 participants where the fit to English nonwords improved if some effect of Spanish probabilities was included. Overall, this effect was positively correlated with Spanish vocabulary (so those with higher Spanish vocabulary were more likely to benefit from mixing in Spanish probabilities). There was also a small negative correlation with English vocabulary (so those with higher English vocabulary were less likely to benefit from mixing in Spanish probabilities). Also, all of those with a high English vocabulary and low Spanish vocabulary did not benefit from mixing Spanish with English to predict English nonwords.

Conclusion

This study provided novel data on well-formedness performance by bilinguals. While the bilinguals, for the most part, performed similarly to monolinguals in each language, some evidence for cross-language effects in well-formedness judgments was found. The presence of cross-language influences appears to reflect individual experience. Individual differences in performance on well-formedness judgments for novel non-words across languages (as well as effects observed within each language) provides strong evidence for a theory of phonotactic knowledge based on emergent constraints derived from individual lexical experience.

References

Coleman, J., & Pierrehumbert, J. B. (1997) Stochastic phonological grammars and acceptability. *3rd Meeting of the ACL Special Interest Group in Computational Phonology: Proceedings of the Workshop, 12 July 1997* (pp. 49-56). Somerset NJ: Association for Computational Linguistics.

Frisch, S. A. (in prep). Vocabulary influences on well-formedness judgments for novel nonwords. MS. University of South Florida.

Frisch, S. A., Large, N. R., & Pisoni, D. B. (2000). Perception of wordlikeness: Effects of segment probability and length on the processing of nonwords. *Journal of Memory and Language*, 42, 481-496.

Hay, J., Pierrehumbert, J. B., & Beckman, M. E. (2004) Speech perception, well-formedness and the statistics of the lexicon. In J. Local, R. Ogden, & R. Temple (Eds.), *Phonetic Interpretation: Papers in Laboratory Phonology VI* (pp. 58-74). Cambridge: Cambridge University Press.

Nusbaum, H.C., Pisoni, D. B., & Davis, C. K. (1984). Sizing up the Hoosier Mental Lexicon: Measuring the familiarity of 20,000 words. *Research on Speech Perception Progress Report No. 10*, 357-376.

Trieman, R., Kessler, B., Knewasser, S., Tincoff, R., & Bowman, M. (2000). English speaker's sensitivity to phonotactic patterns. In M.B. Broe & J.B. Pierrehumbert (Eds.), *Acquisition and the lexicon: Papers in laboratory phonology V* (pp. 369-282). Cambridge: Cambridge University Press.

Perception of Sung and Spoken Vowels in New Zealand English

Andy Gibson

Institute of Culture, Discourse and Communication, AUT University, Auckland, New Zealand;
andy.gibson@aut.ac.nz

This research uses a phoneme boundary experiment to examine the perception of singing and speech in New Zealand. Several previous studies have found that speech perception can be affected by listener expectation. Ladefoged and Broadbent (1957) found that by altering the formant frequencies of an introductory sentence the perception of the following word could also be altered. Hay, Drager, & Warren (2006) found that NZE-speaking participants were better at assigning NEAR and SQUARE[1] words to their lexical sets if the experiment's instructions were given in an RP accent instead of a NZE accent. Recent studies have found that even without preceding linguistic material, listeners' expectations about the social characteristics of a speaker can affect perception (e.g. Niedzielski, 1999; Hay, Nolan, & Drager, 2006; Johnson, Strand, & D'Imperio, 1999; Drager, 2006). These results have been interpreted in an exemplar theory framework by proposing that remembered exemplars of language may be indexed or tagged with relevant social information. The present research will propose that exemplars may also be indexed with salient contextual information. Particularly, it is hypothesized that language which is *sung* may be indexed differently to language which is *spoken*. Most people who listen to popular music are exposed to singing in accents of English different to their own speech. This is especially so for people whose spoken variety is rarely used in singing, such as New Zealanders. The sum of such a person's exposure to a given vowel sound is likely to be systematically different depending on whether the vowel occurred in singing or in speech.

Research conducted into pronunciation in popular singing (e.g. Trudgill 1983, Simpson 1999, Carlsson 2001) has shown that people sing differently than they speak. These singing accents often exhibit features of American English. Coddington (2004) found that even New Zealand artists who expressed a desire to sing in New Zealand English still consistently used a range of American English features in their singing. These singers must navigate a path between a desire to express their own identity (e.g. through the use of their regional dialect), and a compulsion to use the pronunciation most often heard in the genre of music being performed.

If remembered exemplars of sung language are stored differently to exemplars of spoken language, it is likely that in production of singing the sung exemplars are more highly activated and have more influence over the vowel sound to be produced. And so, New Zealanders sing the way they have heard other people sing, and in turn, their listeners build up stronger associations between singing and non-New Zealand English accents. The aim of this research is to test whether New Zealanders' expectations about accents in singing may affect their perception of ambiguous vowel sounds.

The study focuses on the NZE DRESS and TRAP vowels which are raised and fronted compared to other dialects of English. Drager (2006) conducted a phoneme boundary experiment where listeners were asked to decide which of two words they heard. The stimuli were a set of words which ranged on a continuum from sounding like the word *'bed'* to sounding like the word *'bad'*. It was found that the perceived age of the speaker affected the perceived position of the phoneme boundary between the DRESS and TRAP vowels.

The present study tests whether listeners perceive the boundary between DRESS and TRAP differently when they are listening to sung language as opposed to spoken language. The starting assumption is that in the lexicon of a New Zealander, the distribution of *sung exemplars* of DRESS is lower and backer in auditory/acoustic space than the centre of the DRESS exemplar cloud overall. This would be caused by repeated exposure to singing accents with open DRESS vowels and little or no exposure to singing with NZE-like raised DRESS. It is likely that there is considerable acoustic overlap between spoken exemplars of NZE TRAP and sung exemplars of DRESS. This situation could result in listeners perceiving the phoneme

[1] These vowel names (along with DRESS and TRAP below) are based on the lexical sets outlined by Wells (1982).

boundary between the two vowels differently depending on whether they are listening to spoken or sung language.

The methodology is based on that used in Drager (2006). A 12-step continuum was created, with the first two formant frequencies manipulated to create stimuli which sound like 'bed' (F1=400; F2=2150) through to those that sound like 'bad' (F1=765; F2=1820). The continuum was created from one word recorded by the author with formants roughly in the centre of the continuum. Once the continuum was completed, the 'sung' and 'spoken' conditions were created. The 'sung' stimuli were presented to participants in the context of a background musical accompaniment consisting of a recording of guitar, keyboard, bass and drums cycling around the chord progression | G | Em | C | D7 |. In order to give the impression that the voice is actually singing, half of the stimuli in the singing condition were pitch-shifted up one semitone. This difference is small enough not to greatly affect spectral structure but large enough to allow the sung stimuli to be fitted to the musical chord progression. This gives an impression of melodic movement despite the stimuli being relatively fixed in their actual pitch. Each word is presented twice, on the first and third beat of the bar. There is then a gap of one bar for the participant to respond to the task by circling either the word 'bed' or the word 'bad' on a response sheet. In the spoken condition, the stimuli are presented at the same rate but without any background music, and all at the original pitch. The spoken and sung stimuli are presented in separate blocks. In each block, every word from the continuum is played twice in randomized order. This order is always the same in both conditions for any given participant. Half of the participants heard singing first and half heard speech first.

It was hypothesized that in the sung condition a higher percentage of the stimuli would be categorized as 'bed' than in the spoken condition. This hypothesis is based on the idea that a New Zealander's expectation, based on past experience, is to hear lowered DRESS in singing and raised DRESS in speech. These expectations will then affect perception of ambiguous stimuli. Initial results support this hypothesis. Steps 5-8 of the continuum all received a mixture of 'bed' and 'bad' responses. Of all the responses for these ambiguous stimuli, 48% were categorized as 'bed' in the sung condition while only 38% were perceived as 'bed' in the spoken condition. These results provide support for the argument that remembered exemplars of sung language are indexed differently from exemplars of normal speech. Further development of this theory may help to explain why people use different accents in their singing than in their speech.

References

Carlsson, C. J. (2001). The way they sing it: Englishness and pronunciation in English pop and rock. *Moderna Sprak,* 95(2), 161-168.

Coddington, A. (2004). *Singing as we Speak? Pronunciation in New Zealand Popular Music*. M.A. thesis: University of Auckland.

Drager, K. (2006). From Bad to Bed: the relationship between perceived age and vowel perception in New Zealand English. *Te Reo,* 48, 55-68.

Hay, J., Drager, K., & Warren, P. (2006). *Cross-Dialectal Exemplar Priming*. Poster presented at the Tenth Conference on Laboratory Phonology. Paris, June 2006.

Hay, J., Nolan, A., Drager, K. (2006). From fush to feesh: exemplar priming in speech perception. *Linguistic Review,* 23(3), 351–79

Johnson, K., Strand, E., & D'Imperio, M. (1999). Auditory-visual integration of talker gender in vowel perception. *Journal of Phonetics,* 27, 359-84.

Ladefoged, P., & Broadbent, D. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America,* 29(1), 98-104.

Niedzielski N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology,* 18(1), 62–85.

Simpson, P. (1999). Language, culture and identity: With (another) look at accents in pop and rock singing. *Multilingua,* 18(4), 343-367.

Trudgill, P. (1983). *On Dialect: Social and Geographical Perspectives*. Oxford: Blackwell.

Wells, J.C. (1982). *Accents of English*. Cambridge: Cambridge University Press.

Distribution of tongue tip articulations in Hindi versus English and the acquisition of stop place categories

Louis Goldstein*#, Hosung Nam#, Manisha Kulthreshtha#, Leslie Root, and Catherine Best#+
*Department Lingustics, University of Southern California; #Haskins Laboratories, New Haven, CT, USA
+MARCS Auditory Laboratories, University of Western Sydney
louisgol@usc.edu; nam@haskins.yale.edu; manishakulshreshtha@yahoo.com; c.best@uws.edu.au

Explaining the decreased ability of infants around 8-10 months of age to distinguish at least some phonological contrasts absent from their native language (Werker & Tees, 1984) has constituted a theoretical challenge. Since infants of that age do not yet appear to have acquired the phonological system of the ambient language, how is it that the structure of that phonology is nonetheless active in shaping their perceptual behavior? A promising recent solution to this puzzle is provided by Maye et al. (2002), who have argued that the distributional structure of the infant's input (i.e., adult speech), combined with infants' statistical sensitivities, could account for changes in early perception. Supporting this idea, they manipulated the distributional properties of a VOT continuum (voiced to voiceless unaspirated), exposing infants to unimodal vs. bimodal distributions in an experimental session. Bimodal distributions led infants to perceptually discriminate items taken from the distinct modes, whereas they did not discriminate these items when their experience was unimodal.

While this is a promising theoretical account, it begs the questions of how real-world input, which is connected speech, is decomposed by infants into discrete items whose statistical properties can be accumulated and what the relevant dimensions are along which statistics are kept. Potential answers to these questions are provided by a different, though complementary, approach to modeling the infants' developmental course in perception and production, namely, the articulatory organ (AO) hypothesis (Goldstein & Fowler, 2003; Best & McRoberts, 2003). Under this hypothesis, infants can decompose the oral-facial system into distinct organs (e.g., lips vs. tongue tip vs. tongue dorsum) from very early in life (as consistent with neonates' ability to perform facial mimicry: Melzoff & Moore, 1997), and they are sensitive to the actions of these organs in producing constrictions within the vocal tract. This hypothesis predicts that contrasts involving actions of distinct organs (e.g., /b/ vs. /d/) should be mastered relatively early, while those involving quantitatively different actions of the same organ (e.g., /ð/ vs. /d/), should be acquired only when the infant (or child) has attuned sufficiently to the distributional properties of the organ's constrictions. Results from children's early word productions are consistent with this hypothesis (Goldstein, 2003), and several perceptual findings can be explained with reference to it (Best & McRoberts, 2003).

Werker's original experiments demonstrating the loss of ability to discriminate non-native contrasts are clear examples of within-organ contrasts: Hindi dental and retroflex stops are tongue tip constrictions (differing in the location of constriction), and Nthlakapmx velar and uvular stops are tongue dorsum constrictions (also differing in location). Thus loss of ability of English-learning infants to discriminate these stimuli could be explained as a result of their experience with a unimodal distribution of tongue tip or tongue dorsum constrictions, while Hindi- and Nthlakapmx-learning infants would presumably experience bimodal distributions.

Reasonable as this account seems, there are no data to support the hypothesis of bimodality of tongue tip distributions in Hindi (or tongue dorsum distributions in Nthlakapmx). Even though a bimodal distribution might be expected, it is possible that contextual variation in tongue tip positioning (or distributional asymmetries within the language) might obscure an underlying contrast in constriction location, at least in the surface articulation (and resulting sound). We therefore collected data on tongue tip constrictions in running (adult) speech in Hindi, across a range of phonetic contexts, and compared it to data from English.

Method. A female Hindi speaker was recorded reading an approximately 6000-word story, while the positions of her lips, tongue tip, tongue body and jaw were measured using EMMA. Locations of coronal stops were identified from the acoustics, and time during that stop at which the tongue tip receiver was closest to the palate was algorithmically determined. The horizontal position (advanced-retracted) of the tongue tip (TTx) at that maximally constricted time for each stop was logged and the distribution of TTx values was plotted. For comparison, English data from the Wisconsin X-ray database was analyzed for 6 read paragraphs using the same procedures. Because there were fewer coronals in these paragraphs than in the Hindi story, the analysis was carried out for 3 of the X-ray subjects, 2 male and 1 female. The data from each subject (both Hindi and English) was normalized to the range 0 (most advanced) to 1 (most retracted).

Results. As shown in Fig. 1a, the distribution of TTx was bimodal in Hindi. None of the 3 English subjects showed a bimodal distribution. The results pooled across English subjects are shown in Fig. 1b.

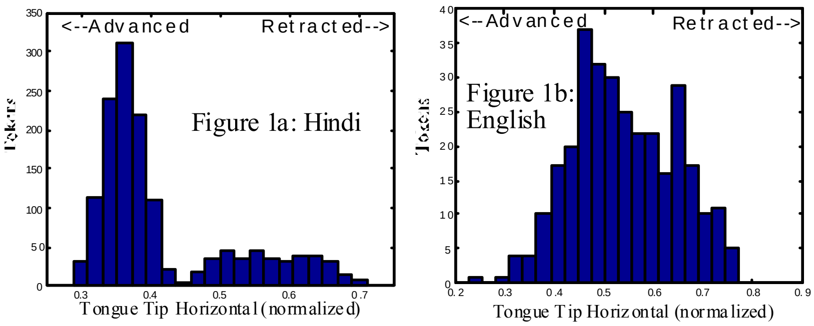


Figure 1: Histogram of tongue tip horizontal positions in (a) Hindi (b) English

While the Hindi distribution *appears* bimodal, the distribution of retroflexes is fairly broad, and it is important to know whether the overall distribution would afford the learning of two distinct constriction categories. To test this, the observed normalized TTx values were input to a Hebbian learning model that has been shown to result in self-organization of discrete phonetic categories from continuous input (Oudeyer, 2006; Nam et al., in press). With the Hindi data as input, the model converged on two sharply distinct categories of neural units, while with the data from any of the English subjects, or with the pooled data, only a single mode developed.

Conclusion. The adult data collected in this study suggest that if infants track distributional properties of tongue tip constrictions within more naturalistic connected speech (however they manage to do that), they will arrive at 2 categories of tongue tip behavior in stops in a Hindi environment, but a single category in an English environment. This, in turn, provides evidence in support of distributional attunement as a possible basis for changes in perceptual discrimination in the case of within-organ contrasts.

Acknowledgments. This research was supported by NIDCD grant DC-00403.

References

Best, C., & McRoberts, G. W. (2003). Infant perception of non-native consonant contrasts that adults assimilate in different ways. *Language & Speech,* 46, 183–216.

Goldstein, L. (2003). Emergence of discrete gestures. In Solé, M.J., Recasens, D., & Romero, J. (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 85-88). Rundle Mall: Causal Productions

Goldstein, L., & Fowler, C. (2003). Articulatory phonology: a phonology for public language use. In Meyer, A. & Schiller, N. (Eds.), *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities* (pp. 159-207). New York: Mouton.

Maye, J., Werker, J., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition,* 82, B101–B111.

Meltzoff, A., & Moore, M. (1997). Explaining facial imitation: a theoretical model. *Early Development and Parenting*, 6, 179-192.

Nam, H., Goldstein, L., & Saltzman, E. (in press). Self-organization of syllable structure: A coupled oscillator model. In Chitoran, I., Coupé,C., Marisco, E., & Pellegrino, F. (Eds.), *Approaches to Phonological Complexity*. Berlin: Mouton.

Oudeyer, P.Y. (2006). *Self-Organization in the Evolution of Speech.* Oxford: Oxford University Press.

Werker, J. & Tees, R. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development,* 7, 49-63.

Phonological, lexical, and frequency factors in coronal stop deletion in early New Zealand English

Gregory Guy*, Jennifer Hay# & Abby Walker+

*Department of Linguistics, New York University, United States of America; gregory.guy@nyu.edu
#Department of Linguistics, University of Canterbury, New Zealand; jen.hay@canterbury.ac.nz
+Department of Linguistics, University of Canterbury, New Zealand; ajw129@student.canterbury.ac.nz

Final coronal stop deletion (CSD) in English – seen in alternations such as *wes(t) side, ol(d) man* – is one of the most-studied phenomena in the field of language variation and change. Previous research has revealed several significant constraints on this process including morphology, following phonological context, and lexical identity (Guy, 1980, 1991; Labov, 1989; Neu, 1980; Santa Ana, 1996). We have conducted a large-scale analysis of this variable in early New Zealand English, incorporating several factors into our model that have not previously been investigated, (e.g., lexical frequency, speech rate), which serve to address issues of grammar and representation.

Using both acoustic and auditory cues, we classified over 4000 tokens of underlying t/d as either present or absent. These tokens were drawn from 19 speakers of early New Zealand English. We fit a logistic regression model to the data by hand. The Wald Statistics for the resulting model are shown in Table 1. As the focus of this paper is on linguistic constraints on CSD, we include speaker identity in our model, so as to hold social factors constant.

Table 1: Wald Statistics for model of t/d deletion

Log local frequency	12.2	1	0.0005
t vs d	16.56	1	<.0001
following environment	623.48	5	<.0001
most common following environment	23.08	5	0.0003
and (Factor+Higher Order Factors)	86.94	2	<.0001
hesitation (Factor+Higher Order Factors)	59.55	2	<.0001
Speaker	190.75	18	<.0001
morphology	7.62	2	0.0221
went	17.86	1	<.0001
not	19.22	1	<.0001
local speech rate	5.27	1	0.0217
and * hesitation (Factor+Higher Order Factors)	13.42	1	0.0002
TOTAL	1114.88	38	<.0001

The effect of lexical frequency is highly significant: more frequent words are more often reduced. Interestingly, this factor appears to predict much of the morphology effect that has received so much attention in prior studies, due to the differing frequency profiles of the relevant morphological categories. Underived lexical items (e.g., *pact, mist*) robustly undergo more deletion than regular past tense forms (e.g., *packed, missed*), with irregular past tense forms (*kept, left*) showing intermediate deletion rates. These results are replicated in our data, but underived forms tend to have much higher frequencies than past tense forms, and irregular past tense forms have higher frequencies than regulars. Once frequency is taken into account, a much more modest role for morphology remains.

Notably, the frequency effect is captured using frequency counts taken over the corpus on which our analysis is based – recordings of early New Zealand English collected in the 1940s. While these counts are highly correlated with counts in a larger modern corpus (CELEX), the larger corpus does not significantly predict CSD in our data, due perhaps to dialect differences (CELEX is not based on NZ English), or to significant language change in the 60 years since our data were recorded. In any case, this points to the need for appropriate sources of frequency data that reflect the speakers' own experience.

Another lexical property is the exceptional nature of several frequent items. Thus the word *and* occurs with a higher rate of deletion than is expected on the basis of frequency alone.The same is true of the cliticized negative *-n't* (e.g., *wasn't, didn't*). Interestingly, *and* interacts with whether the following word is a hesitation: phrases like '*and uh*' show much lower rates of deletion, pointing to storage of these highly frequent collocations.

Our results replicate those of previous studies with respect to the strong effect of following context, but we also find that the most common following context for a word has a significant effect. If a given word often occurs in a deletion-favouring environment, then deletion rates on that word are high – *even when that environment is not present.*

Finally, there is an effect of speech rate. We crudely calculated the speech rate in words/second in a 20 second window surrounding the token. We found that faster speech rates lead to greater reduction. An effect of speech rate has been observed on deletion of /t/ and /d/ in other environments (Raymond, Dautricourt & Hume, 2006).

Many of these results are consistent with the predictions of exemplar theory. The frequency and lexical effects, and the effect of an absent most common following context, are all predicted by a mental representation of remembered exemplars. A conventional abstract phonological model can straightforwardly handle some of these results, such as the remaining morphological effect, and following context, but would require highly enriched lexical representations to accommodate the frequency and lexical effects (for example, the exceptional *and* cases might be modeled with multiple underlying representations *an, and, and uh*). Such 'enriched' mental representations, are of course, the basic insight of an exemplar model.

References

Guy, Gregory R. (1980). Variation in the group and the individual: the case of final stop deletion. In W. Labov (Ed.), *Locating language in time and space* (pp. 1-36). New York: Academic Press

Guy, Gregory R. (1991). Explanation in variable lexical phonology: An exponential model of morphological constraints. *Language Variation and Change*, 3, 1-22.

Labov, William (1989). The child as linguistic historian. *Language Variation and Change*, 1, 85-97.

Neu, Helene (1980). Ranking of constraints on /t,d/ deletion in American English. In W. Labov (Ed.), *Locating language in time and space* (pp. 37-54). New York: Academic Press.

Santa Ana, Otto (1996). Sonority and syllable structure in Chicano English. *Language Variation and Change*, 8, 63-89.

Raymond, William, Dautricourt, Robin, & Hume, Elizabeth (2006). Word medial /t, d/ deletion in spontaneous speech: modeling the effects of extra-linguistics, lexical, and phonological factors. *Language Variation and Change*, 18, 55-97

Phonological Relationships: Beyond Contrast and Allophony

Kathleen Currie Hall

Department of Linguistics, The Ohio State University, USA; kchall@ling.osu.edu

In the standard view of phonology, there are two basic paradigmatic relationships between phonological segments: contrast and allophony (see, e.g., Steriade 2007). Two segments are assumed to be contrastive if and only if their distribution in the lexicon of a language is not predictable: if in *at least one* phonological context, it is not possible to predict which of two sounds will occur, those two sounds are contrastive. Allophony is defined as the opposite of contrast: two sounds are phonologically allophonic if and only if their distribution is predictable in *all* phonological contexts. There is increasing evidence, however, that these two types of relationships are insufficient to capture all of the possible relations between segments that are important to language users (see, e.g., Ladd 2006). It is proposed here that phonological relationships should be characterized by a *continuum* of predictability, rather than a categorical distinction between two types of predictability. This proposal accounts for many of the seemingly "problematic" cases of contrast and allophony that have been reported, and has a number of implications for our understanding of both synchronic and diachronic phonology.

The difference between the traditional approach to contrast and allophony and the approach proposed here is illustrated in (1). Each circle represents a phonological environment that a segment can appear in, such as "word finally." The black square in each circle represents one realization of a phonological category, such as [t] or [t^h] or [d]. Sounds such as [t] and [t^h] in English occur in environments that do not overlap at all, and are thus allophonic; sounds such as [t^h] and [d] in English occur in many overlapping environments and are therefore contrastive. As Goldsmith (1995) points out, the traditional divide between these two relationships is entirely uneven: allophony accounts for *only* cases of 100% predictability, while contrast accounts for cases from zero predictability to nearly complete predictability. Goldsmith suggests that it may be useful to reconsider this division, instead conceptualizing these relationships as a "cline," as illustrated by the partially overlapping circles between the endpoints of contrast and allophony in (1). In this paper, the evidence for such a reconsideration is brought forth and the consequences of the new analysis are discussed.

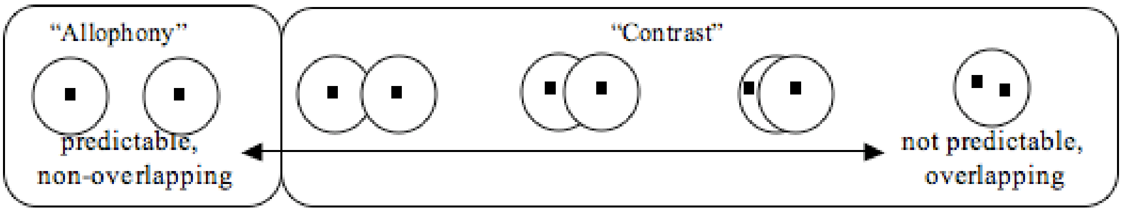


Figure 1: Typical division of the proposed continuum of predictability.

Evidence for treating phonological relationships as a continuum comes from a number of different sources. For example, it is well known that two contrastive segments in a language can also be predictable in certain environments: this is the phenomenon of neutralization, and is traditionally treated as the result of one or more rules that eliminate contrastive distinctions in certain environments. In Optimality Theory, however, neutralization and allophony are different outcomes of the same phenomenon. "Neutralization" is encoded by ranking markedness constraints above faithfulness constraints that would keep a lexical contrast intact. The same mechanism is used to encode "allophony": markedness constraints force potential lexical contrasts into predictable relationships. There is no *a priori* reason to think that contrast neutralization differs from allophony: both involve greater predictability of the distribution of segments than does contrast.

In addition to this theoretical argument, there is evidence that language users *perceive* more phonological relationships than just contrast and allophony. Boomershine et al. (to appear) showed the basic fact that two contrastive segments are perceptually more distinct than two allophonic segments. Munson et al. (2008) showed that, in addition, the neutralization of segments can lead to a distinct perceptual relationship. In

Japanese, the fricatives [s] and [ɕ] contrast before the vowels [a, o, u], but are neutralized to [ɕ] before [i]. Munson et al. demonstrated that Japanese listeners perceived tokens of [ɕ] and [s] before [a, o, u] as distinct tokens of [ɕ] and [s], but perceived tokens of [ɕi] as more [s]-like than [ɕ]-like. By comparison, English listeners, for whom /s/ and /ɕ/ (mapped to /ʃ/) contrast before [a, o, u, i], did not perceive any differences between the [ɕ] tokens before any of the vowels, indicating that the distribution of fricatives in Japanese led Japanese listeners to perceive neutralized contrasts as more similar than non-neutralized ones.

Consider now how redefining phonological relationships to reflect a continuum of predictability can help explain the long-standing debate about the status of [ɑi] and [ʌi] in Canadian English (see, e.g., Mielke et al. 2003). Generally, these vowels are predictably distributed: [ʌi] occurs before tautosyllabic voiceless segments, while [ɑi] occurs elsewhere. However, there are also minimal pairs such as *writer* [ɹʌiɾɹ̩] and *rider* [ɹɑiɾɹ̩] where the two vowels seem to contrast before [ɾ]. Analyzing this distribution as "contrastive" is unsatisfactory because it misses the generalization that the distribution *is* largely predictable; but maintaining the assumption that it is "allophonic" is problematic given that there are in fact minimal pairs. Recasting the problem in terms of a predictability distribution resolves the issue. Of the approximately 66 different following segments that [ɑi] and [ʌi] can appear before, only one ("before [ɾ]") shows unpredictability—that is, [ɑi] and [ʌi] are predictable in 98.5% of environments, and not in 1.5%.

This intermediate status between contrast and allophony is just that—intermediate. There is no need to force the distribution to either end of the continuum of predictability; for example, Ernestus (2006) has shown that language users are sensitive to probabilistic distributions of segments, regardless of whether these distributions are categorically labelled "contrastive" or "allophonic" by linguists. This intermediate status can also contribute to our understanding of diachronic changes such as "phonemic splits"—instances where, despite being in predictable distributions, two segments have been treated by speakers as being contrastive, as evidenced by a subsequent "split" into separate and unpredictable distributions. In Canadian English, for example, the traditional allophonic distribution is beginning to break down, even in non-[ɾ] environments, and [ɑi] and [ʌi] can sometimes occur in unpredicted environments (e.g., [ɑi] appearing in *like*, or [ʌi] appearing in *gigantic*; see Hall 2005). This split is a logical consequence of a situation where the vowels are predictably distributed in some, but not all, of their environments. Thus, reanalyzing phonological relationships to reflect their probabilistic nature elucidates many of the apparent anomalies arising from traditional definitions of contrast and allophony in both synchronic and diachronic phonology.

References

Boomershine, A., Hall, K. C., Hume, E., & Johnson, K. (to appear). The influence of allophony vs. Contrast on perception: The case of Spanish and English. In P. Avery, B. E. Dresher & K. Rice (Eds.), *Contrast in phonology: Perception and acquisition*. Berlin: Mouton.

Ernestus, M. (2006). Statistically gradient generalizations for contrastive phonological features. *The Linguistic Review,* 23, 217-233.

Goldsmith, J. (1995). Phonological theory. In J. A. Goldsmith (Ed.), *The Handbook of Phonological Theory* (pp. 1-23). Cambridge, MA: Blackwell.

Hall, K. C. (2005). Defining phonological rules over lexical neighbourhoods: Evidence from Canadian raising. In J. Alderete, C.-h. Han & A. Kochetov (Eds.), *Proceedings of the 24th West Coast Conference on Formal Linguistics* (pp. 191-199). Somerville, MA: Cascadilla Proceedings Project.

Ladd, D. R. (2006). "Distinctive phones" in surface representation. In L. M. Goldstein, D. H. Whalen & C. T. Best (Eds.), *Laboratory Phonology 8* (pp. 3-26). Berlin: Mouton de Gruyter.

Mielke, J., Armstrong, M., & Hume, E. (2003). Looking through opacity. *Theoretical linguistics,* 29, 123-139.

Munson, B., Li, F., Yoneyama, K., Hall, K. C., & Beckman, M. (2008). *Sibilant fricatives in Japanese and English: Different in production or perception?* Paper presented at the Linguistic Society of America, Chicago, IL.

Steriade, D. (2007). Contrast. In P. de Lacy (Ed.), *The Cambridge Handbook of Phonology* (pp. 139-157). Cambridge: Cambridge University Press.

When Ongoing Historical Sound Change Meets with Social Factors

Fang Hu

Phonetics Lab, Institute of Linguistics, Chinese Academy of Social Sciences, China; hufang@cass.org.cn

Introduction

Language is always on the way of change from a variationist perspective. Diachronically, the phonetics and phonology of a language result from historical processes; synchronically, the lexicon is subject to social and idiosyncratic variations. In this connection, the *present* regional, social, and idiosyncratic variations across speech communities in a kindred language group are of central interest to historical linguistics as well as sociophonetics, since they pursue the facts of sound change and the mechanisms of sound variations and thus probe into the *past* and even predict the *future* of the phonetics and phonology of that language group (Labov, 1994, 2001; Foulkes & Docherty, 2006). In this paper, I will discuss a particular case where the *internal* historical sound change meets with *external* social factors, namely how the post-oralization of the nasal initial consonants in Chinese dialects groups (Hu, 2007) are influenced by the authoritative Standard Chinese.

Methodology

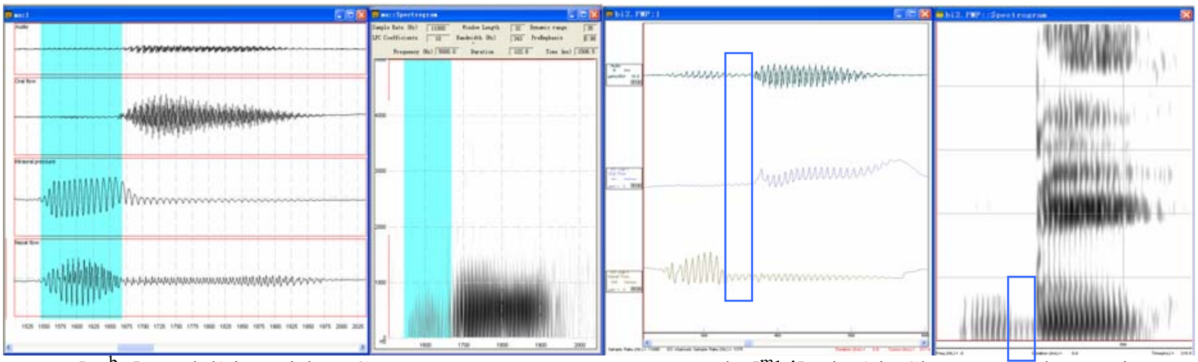
Four major Chinese dialect groups were investigated. For Shanxi Jin, 51 native speakers from 39 counties were recorded acoustically and among them, 28 speakers from 19 counties were additionally recorded aerodynamically using PCQuirer. For Cantonese, six speakers from the Zhongshan area were recorded aerodynamically. For southern Min, six native male speakers from the Chao-Shan area were recorded aerodynamically. For Hakka, one male speaker from Qingxin was recorded aerodynamically. In addition to the data obtained from my own fieldwork, published audio data were also used in the present study, namely the Xiamen, Shantou, and Pingyao data from *Hanyu Fanyan Yinku* ‘Sound compilation of Chinese dialects’.

Speech material was designed for each dialect group. Natural monosyllabic words in isolation were used as test words, which contain all the target nasal consonants, regardless of plain or post-oralized, and the corresponding stops or affricates/fricatives for the comparison’s sake. The speech material was balanced concerning the vowel category and syllable structure. Three repetitions were recorded for each test word.

Post-oralization of nasal initials in Chinese dialects

Post-oralization of nasal initial consonants is widely detected in Chinese dialects (Chan, 1987; Hu, 2007). The plain nasal initials in Middle Chinese may remain intact as in Wu Chinese, have conditionally changed into plain fricatives or approximants as in Mandarin Chinese, or become post-oralized. Triggered by the orality on the following vowel, the post-oralization of nasal initials in Chinese dialects was and is proceeding in a gradient, rather than a categorical way. Cantonese dialects around the Zhongshan area, the Qingxin Hakka, and the Shanxi Jin dialects are at an early stage of post-oralization, where the nasal initials acquire orality only at release; that is, they are orally released (poststopped or post-fricated/affticated) nasals. The southern Min dialects are at the late stage, where the orality has spread from the release to the closure portion of the nasal consonants; in other words, the post-oralized nasals have become canonical prenasalized consonants. Figure 1 compares typical productions for the orally released nasal (1a) and the prenasalized consonant (1b). As shown by the shaded part in Fig. 1a, the nasal airflow for [m^bu] ‘model’ in Taishan Cantonese continues until the oral release begins; as indicated by the superimposed rectangle in Fig. 1b, there is an interval between the diminution of nasal airflow and stop release for [^{m}bi] ‘rice’ in Shantou southern Min, suggesting a transition from a nasal to an oral stop.

Post-oralization in Chinese dialects begins as a phonetic process, namely the post-oralized nasal consonants usually do not contrast with plain nasal consonants in most cases. For instance, the post-oralized nasals occur before oral vowels while the plain nasal consonants before nasalized vowels in Xiamen southern Min; poststopping of the nasals is a phonetic variation in the Cantonese dialects. However, in some dialects, post-oralization has become a phonological process, namely phonological contrasts occur. For instance, [^{m}bak] ‘wood’ vs. [mak] ‘eyes’ in Shantou southern Min; [n^{z}ɑŋ] ‘to speak to oneself’ vs. [n^{d}ɑŋ] ‘south’ in Wenshui and Pingyao Jin.



a. [m^{b}u]‘model’ in Taishan Cantonese. b. [^{m}bi] ‘rice’ in Shantou southern Min

Figure 1: The audio, oral flow, intraoral pressure (available in 1a only) and nasal flow (left), and the wideband spectrogram (right) for the production of an orally released nasal (1a) and a prenasalized consonant (1b).

Social factors

The ongoing post-oralization in Chinese dialects is an internal process of sound change in the sense that it is triggered by the linguistic-internal factor. Meanwhile, the post-oralization is subject to external social factors. What is central is that the speakers, especially those younger in the speech communities, are influenced greatly by the authoritative Standard Chinese from education, the media, and daily communications with non-local people. In Standard Chinese, there is no post-oralization present at its phonetic variations. Both the two nasal initials [m n] and the two nasal codas [n ŋ] are plain nasals. And the historical alveolo-palatal nasal initial has changed into a post-alveolar approximant, which is an internalized phonological knowledge of the speakers.

It is quite interesting to observe that the post-oralization at different stages exhibits totally different patterns when meeting with the same social factors. As in a late stage of post-oralization, the prenasalized consonants in southern Min exhibit a strong tendency to be developing to the corresponding plain voiced oral consonants. But the post-oralization at early stage tends to hinge on social impacts in Cantonese, Hakka and Shanxi Jin. That is, the orally released nasal consonants in these dialects show a tendency to change back to plain nasal consonants. However, it is intriguing to note that the internal linguistic factors interact with external factors in a quite complicated way. The poststopped nasals are more likely to change back to plain nasals, since the variation is at a phonetic level; the post-fricated/affricated nasals show more resistibility to social factors, as the variation is at a phonological level, i.e. internalized in the speaker’s lexicon.

References

Chan, M. (1987). Post-stopped nasals in Chinese: An areal study. *UCLA Working Papers in Phonetics*, 68, 73-119.

Foulkes, P., & Docherty, G. (2006). The social life of phonetics and phonology. *Journal of Phonetics*, 34, 405-530.

Hu, F. (2007). Post-oralized nasal consonants in Chinese dialects. In Jürgen Trouvain & William J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 1405-8). Saarbrücken, Germany.

Labov, W. (1994). *Principles of Linguistic Change (Volume 1): Internal Factors*. Oxford: Blackwell.

Labov, W. (2001). *Principles of Linguistic Change (Volume 2): Social Factors*. Oxford: Blackwell.

Experiments on Fine Phonetic Detail of Arabic Syllable Structure

Michael Ingleby & Azra N. Ali

School of Computing and Engineering, University of Huddersfield, England;
m.ingleby@hud.ac.uk a.n.ali@hud.ac.uk

Empirical investigation of fine consonant contrasts in different dialects of Arabic have included timing approaches to contrast the duration of the audio signal of a single consonant (singleton) with that of word-internal geminates in the same word context (Hassan, 2002; Ingleby & Boathman, 2002; Khattab, 2007). These studies have shown that the duration of geminates is about twice that of a singleton (Figure 1a, following page). Such studies are concerned with external <u>articulatory</u> contrasts, but in these experiments we investigate <u>auditory and cognitive</u> aspects of the contrasts, probing the mental models that mediate perception. Given incongruent stimuli, in which the audio and visual signals are in conflict, many observers report a perception that differs from the data in either channel. The phenomenon is known as McGurk fusion (McGurk & MacDonald, 1976). Typically audio 'ba' temporally aligned with visual 'ga' (lip movements) elicits the fusion percept 'da', or more symbolically

(1) $_{AUD}(\textbf{\textit{b}}\textit{ait} \,\|\, \textbf{\textit{g}}\textit{ate})_{VIS} \rightarrow (\textbf{\textit{d}}\textit{ate})_{PER}$
and (2) $_{AUD}(\textit{ma}\textbf{\textit{p}} \,\|\, \textit{ma}\textbf{\textit{ck}})_{VIS} \rightarrow (\textit{ma}\textbf{\textit{t}})_{PER}$
and (3) $_{AUD}(\textbf{\textit{b}}\textit{aal} \,\|\, \textbf{\textit{q}}\textit{aal})_{VIS} \rightarrow (\textbf{\textit{d}}\textit{aal})_{PER}$
and also (4) $_{AUD}(\textit{naha}\textbf{\textit{b}} \,\|\, \textit{naha}\textbf{\textit{q}})_{VIS} \rightarrow (\textit{naha}\textbf{\textit{d}})_{PER}$

For an incongruent phonetic segment, the <u>fusion rate</u> (proportion of participants reporting fusion depends on the segment's syllabic context, and can serve as a probe of structure. When English syllable structure is probed in this way, the abiding pattern is that for fusion rates amongst *Anglophones* are significantly less for syllabic onsets than for codas (Ali & Ingleby, 2005). The onset/coda differences survive in branching constituents, polysyllabic words and words embedded in natural phrase contexts. The onset/coda differences were, however, <u>not observed</u> when Arabic stimuli were put to *Arabophones* - thus adding to growing evidence that Arabic may be a coda-less CV language at the level of the mechanisms mediating perception in *Arabophones* (Ali et al., 2005).

The experiments in this paper move further towards isolating differences in cognition of fine phonetic detail between *Anglophones* and *Arabophones*. We use Arabic stimuli with word-internal audiovisual incongruity at sites of singleton/geminate phonemic contrast, as exemplified by

(5) $_{AUD}(\text{ħabaa} \,\|\, \text{ħaqaa})_{VIS} \rightarrow (\text{ħadaa})_{PER}$ *vs* $_{AUD}(\text{ħabb-a} \,\|\, \text{ħaqq-a})_{VIS} \rightarrow (\text{ħadd-a})_{PER}$

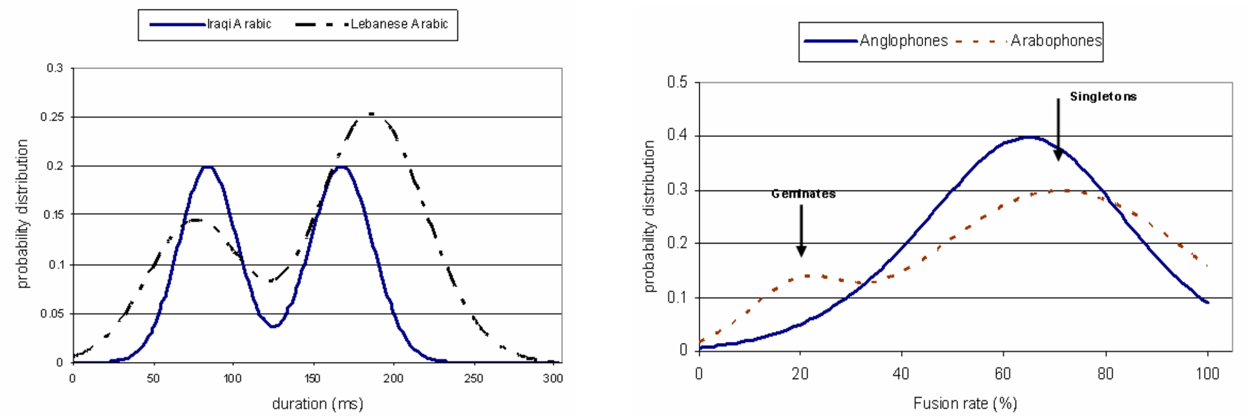
In the first experiment, we put the Arabic stimuli (incongruent stimuli randomized amongst distracting congruent stimuli) to *Arabophones*. In the second experiment we put the Arabic words to *Anglophones*. In both experiments participants were given open choice response forms; in Arabic script for *Arabophones* (e.g. حَبّ) and transcription form for *Anglophones* (e.g. 'habba'). None of the participants had any linguistic knowledge and were allowed to replay the video clip at will before reporting a percept.

The results showed distribution of fusion rates is bimodal for Arabophones, with a lower-rate peak for geminate consonants and a higher peak for singletons. But, for *Anglophones* there was a unimodal distribution law that indicates a failure to perceive the singleton/geminate contrast. We impute this difference of perception between our two subgroups of participants to a difference of mental models, similar to the differences of model reflected in the failure of *Arabophones* to perceive onset/coda contrasts.

Such differences of mental model internalize obvious habituation patterns. Externally, singleton/geminate contrast is semantically important and very common in Arabic; but in English, though textual gemination is a common orthographic feature, truly phonemic gemination is rarer. When attested, it is a product of collisions – morphological (e.g. 'u**nkn**own' and 'sou**ll**ess'), or cross-word (e.g 'bi**g g**ame' or 'to**p p**ost') or across phrase boundaries (e.g. 'Ja**ck, c**utting in, said…' or ('Po**p, p**osing a question, stood…'). Durational studies (Benus, Smorodinksy & Gafos, 2004) show that these lead to corresponding gradience in geminate binding: shortest duration, strongest binding for morphological collisions; longer duration, weaker binding at word boundaries; weakest binding at phrase boundaries, where the germinate integrity may even be compromised by a reduplicated burst. The collision phenomena of Arabic are made more complex by case endings, glottal stops and normative elision processes in possessives, so for present purposes we have investigated only word-internal cases. They are internalized differently by *Arabophones* and *Anglophones*.

The differences of perceiving coda/onset fine contrasts may also be internalizations, but of the more elusive notion of syllable. *Arabophones* are exposed to CV patterns through their orthography in which words are built from consonants pointed with vowel diacritics, and through traditions of Qur'anic recitation. This has been represented in a CV element phonology for Arabic recently (Ingleby & Baothman, 2002). The use of codas (closed syllables) in the teaching of Arabic prosody is part of a Western classical tradition borrowed from Latin and Greek verse and used in studies of prosody in Germanic and Romance languages.

Whatever habituation patterns are responsible for the observations, there is a need to probe further: larger samples, corroborative priming experiments, the usual procedures of psycholinguistics. And in the longer term, a comparative psycholinguistic study of collision phenomena and cross-boundary coarticulation will be needed before internalization in the mental lexicon is fully understood.



(a) (b)

Figure 1: Schematic distribution of (a) duration of singleton and geminate consonants (b) fusion rates for singleton and geminate consonants.

References

Ali, A.N., & Ingleby, M. (2005). Probing the syllabic structure of words using the audio-visual McGurk Effect. *Proceedings of XXVII Annual Conference of the Cognitive Science Society* (pp.91-7). Stresa, Italy.

Ali, A.N., Hassan-Haj, A., Ingleby, M., & Idrissi, A (2005). McGurk fusion effects in Arabic words. In Eric Vatikiotis-Bateson (Ed.) *Auditory-Visual Speech Processing 2005* (pp.11-16). Vancouver.

Benus, S., Smorodinsky, I., & Gafos, A. (2004). Gestural coordination and the distribution of English "geminates". In S. Arunachalam & T. Scheffler (Eds.), *Proceedings of 27th Penn Linguistic Colloquium* (pp.33-46). Penn Linguistics Club.

Hassan, Z. M. (2002) Gemination in Swedish and Arabic with a particular reference to the preceding vowel duration. An instrumental and comparative approach. *TMH-QPSR* Vol.44, Fonetik.

Ingleby, M., & Baothman, F (2002). Empty nuclei in Arabic speech patterns and the diacritic sukuun. In Sami Boudelaa (Ed.) *Perspectives on Arabic Linguistics XVI.* (pp83-94). Amsterdam: John Benjamin.

Khattab, G. (2007) A phonetic study of gemination in Lebanese Arabic. In Jürgen Trouvain & William J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences* (pp153-8). Saarbrücken.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264, 746-748.

Listeners Use Vowel Harmony and Word-Final Stress to Spot Nonsense Words: A Study of Turkish and French

Barış Kabak*, Kazumi Maniwa* & Nina Kazanina#

*Department of Linguistics, University of Konstanz, Germany;
baris.kabak@uni-konstanz.de, kazumi.maniwa@uni-konstanz.de
#Department of Experimental Psychology, University of Bristol, UK; nina.kazanina@bristol.ac.uk

Speakers' knowledge of sound distributions and rhythmic alternations that systematically characterize wordhood in individual languages is known to aid word segmentation. Vowel harmony is one such regularity that dictates a set of co-occurrence restrictions on vowel features within a word, e.g. in Finnish and Turkish, all vowels within a word must agree on the front-back dimension. In these languages opposite values of the front/back feature on adjacent vowels automatically signals a word boundary, as disharmony is not expected within single words. Accordingly, Finnish speakers detect target words faster when the preceding syllable contains a vowel that differs on the front-back dimension from the vowels in the target (Suomi et al., 1997). Likewise, the culminative nature of accent, which requires that every lexical word has one primary stress, is also known to aid speech segmentation. Especially when primary stress is fixed to a particular position that demarcates word boundaries, as in word initial- or word final-stress languages, this may provide the language user with invaluable cues to detect word boundaries. This idea found support in a previous study which reported a facilitatory effect of word-initial stress in Finnish (Vroomen et al., 1998). Since primary stress overlaps with the beginning of words in this language, it is difficult to know whether facilitation effects are due to (i) the demarcative function of stress per se, which prompts a word boundary before the stressed syllable, or (ii) the well-known primacy of word onsets in general. Instead, we test Turkish and French, where stress typically falls on the word-final syllable, and thus separate the demarcative function of stress from the primacy of word onsets. We demonstrate that listeners employ word-final stress cues to progressively postulate an upcoming word boundary. Furthermore, we show that detection of a vowel harmony mismatch, which unlike word-final stress constitutes a regressively operating cue for a word boundary, is robustly exploited only by Turkish listeners. This finds a straightforward explanation since Turkish, but not French, has front-back vowel harmony. Thus, we show that listeners can exploit abstract phonological regularities in their native language to segment even nonsense words.

We conducted a target-detection task that employed a 2x2x2 design with the factors *language* (Turkish/French), *stress* (stress2/stress3) and *harmony* (match/mismatch). Participants heard a 5-syllable CVCVCVCVCV auditory string that consisted of a trisyllabic pre-target string and a disyllabic target (Table 1). The pre-target string and the target were both nonwords in Turkish and French, and were harmonious, i.e. each contained only front or only back vowels. However, in half of the cases the pre-target and the target matched on the frontness/backness dimension, their concatenation contained only front vowels or only back vowels (the harmony-match conditions). In the remaining cases, the pre-target contained front vowels and the target contained back vowels or vice versa (harmony-mismatch). Furthermore, the location of stress in the pre-target was manipulated so that it fell either on the 2nd/3rd syllable (stress2 vs. stress3 conditions). On each trial, the participants were prompted with a visual target, e.g. *pαvo*, which was then followed by an *auditory* 5- syllable nonsense string, e.g. *golu**sho**pαvo*. The task was to determine whether the auditory string contained the visual prompt as quickly and accurately as possible (the correct response was always 'Yes' for experimental items).

Table 1. A sample set of conditions for the target *pαvo*. The stressed syllables (in bold) are longer than the unstressed syllables (240 vs. 160 ms), in their F0 range and F0 contour. Front vowels are in grey and back vowels are in black.

	harmony-match	**harmony-mismatch**
stress 2	go**lu**shopα**vo**	go**lu**shopα**vo**
stress 3	golu**sho**pα**vo**	golu**sho**pα**vo**

Response times (RTs) were measured from the onset of the target in each auditory string. Thirty two sets of experimental materials were distributed across 4 presentation lists following a Latin Square design. Each list also contained 224 filler items to ensure an equal proportion of 'Yes/No' responses across all items, an equal number of harmonic/disharmonic targets and an equal probability of a target word occurring in different positions within an auditory string. Given that stress in both Turkish and French signals a word-boundary immediately after the stressed syllable, identifying the target nonword should be easier in the stress3 conditions than in the stress2 conditions in both languages. In addition, targets should be detected faster and/or more accurately in the harmony-mismatch conditions than in the harmony-match conditions in Turkish, but not in French. Mean accuracy rates and RTs for experimental items based on 40 Turkish and 40 French speakers are summarized in Table 2. RTs below 300 ms and those that exceeded a threshold of 2.5 standard deviations above a participant's mean reading rate for experimental items were replaced by the threshold value; incorrectly responded trials were excluded from the RT analyses. Consequently, a number of conditions in some sets were left with no data points, hence the corresponding sets had to be excluded in order to preserve the validity of the items analysis (1 set excluded in French, 6 sets in Turkish).

Table 2	**Turkish (n=40)**		**French (n=40)**	
	% correct	*RT (st.err.)*	*% correct*	*RT (st.err.)*
stress 2, match	69.7 (2.6)	895 (25)	84.1 (2.1)	950 (20)
stress 2, mismatch	88.1 (1.8)	872 (21)	89.1 (1.7)	948 (21)
stress 3, match	72.2 (2.5)	772 (19)	87.8 (1.8)	914 (23)
stress 3, mismatch	94.4 (1.3)	733 (19)	90.6 (1.6)	831 (21)

Accuracy: There was no difference in accuracy rates to filler items between the Turkish and the French groups (Turkish = 85.5%, French = 86.6%). In 2x2x2 ANOVAs on experimental items, main effects of *language*, *stress,* and *harmony* were all significant and, critically, the interaction *language* x *harmony* was significant. 2x2 ANOVAs within each language group revealed a robust significant effect of harmony in the Turkish group due to higher accuracy rates in the harmony-mismatch conditions than in the harmony-match conditions (91.2 vs. 70.9%). No similar robust effect of harmony was found in the French group (89.8 vs. 85.9%). **RTs:** 2x2x2 ANOVAs showed a marginally significant *language* x *stress* interaction and *language* x *stress* x *harmony* interaction. 2x2 ANOVAs within each language group revealed a significant main effect of stress in both language groups. In French, the interaction *stress .harmony* was also significant. Post-hoc analyses (with Bonferroni correction) showed that the effect of stress was significant both in the harmony-match and harmony-mismatch conditions in Turkish, whereas in French this was significant only in the harmony-mismatch conditions but not in the harmony-match conditions.

Manipulation of the position of stress yielded a significant effect on RTs in both languages. Conversely, harmony had a robust effect only on accuracy rates in Turkish. These results support the claim that stress and vowel harmony regularities that bear demarcative functions can facilitate speech segmentation, and this also applies to non-words. They also suggest that while speakers of languages with a fixed stress may be stress-deaf, i.e. unable to robustly identify the location of the stress in the word or even discriminate two words on the basis of a differential location of stress (Dupoux et al., 1997), they can successfully use the same cue for word segmentation. This is on a par with allophonic and durational regularities which have been shown to be exploited by speakers in word recognition tasks but are not substantially and consistently operationalized for identification or discrimination purposes in speech perception tasks (e.g., Whalen, Best & Irwin, 1997).

References

Dupoux, E., Pallier, C., Sebastián-Gallés, N., & Mehler, J. (1997). A destressing deafness in French. *Journal of Memory and Language*, 36, 399-421.

Suomi, K., McQueen, J. M., & Cutler, A. (1997). Vowel harmony and speech segmentation in Finnish. *Journal of Memory & Language*, 36, 422-444.

Vroomen, J., Tuomainen, J., & de Gelder, B. (1998). The roles of word stress and vowel harmony in speech segmentation. *Journal of Memory and Language*, 38*,* 133-149.

Whalen, D. H., Best, C. T., & Irwin, J. (1997). Lexical effects in the perception and production of American English /p/ allophones. *Journal of Phonetics,* 25, 501-528.

Influence of Prosodic Phrasing on Stop Voicing and Vowel Hiatus Resolution in Modern Greek

Evia Kainada

Department of Linguistics and English Language, University of Edinburgh, Scotland;
evia@ling.ed.ac.uk

It is well established that the phonetic realisation of segments is influenced by many factors, one of which is the position of the segment in the prosodic structure of the language (e.g. Fougeron & Keating 1997). The present study investigates the effect of prosodic boundary strength on two sandhi processes in Greek, namely resolution of vowel hiatus and postnasal stop voicing.

Previous research takes for granted that the application of phenomena such as resolution of vowel hiatus is categorical. For example, in Modern Greek in the context of [to # a'maksi]'the car', full V1 deletion has been expected, yielding [ta'maksi] (Kaisse 1985, Casali 1997). Based on such assumptions regarding the application of sandhi processes, researchers postulated the existence of prosodic constituents within a prosodic structure (e.g. Nespor and Vogel 1986). However, recent research demonstrates that several processes are influenced by the prosodic hierarchy in a gradient manner (e.g. Wightman et al. 1992 for pre-boundary durations, Fougeron & Keating 1997 for post-boundary strengthening). In particular regarding sandhi phenomena, recent studies have shown that processes like vowel hiatus have a phonetically non-categorical output within domains (Zsiga 1997, Baltazani 2006) ranging from varying degrees of coarticulation to total deletion of one of the two vowels in hiatus. Such findings have left researchers with the question of which processes are restricted to applying within single domains (like the sandhi analyses proposed by Nespor & Vogel 1986), and which reflect the hierarchical nature of prosodic constituent structure in applying gradiently across domains (like vowel hiatus, and pre- and post-boundary durations).

This paper deals with the effect of prosodic boundary strength on two segmental sandhi processes. We test whether there is a prosodic effect on the phonetic output of the segments, and - more importantly - if the effect of prosodic structure operates in similar ways on these two processes.

To address these issues, production data were collected from five native speakers of Modern Greek (two male and three female). The materials consisted of an *Adjective* (W1) followed by a *Noun* (W2). The following table shows how the materials were constructed for each process along with an example.

	W1 ending in	W2 starting with		example
Vowel Hiatus	[-ta] or [-to]	[apo-] or [opo-] or [omo-]	hiatus sequences: [ao], [oa] and [aa], [oo] as baselines	[a'notato a'poktima] 'supreme asset'
Stop Voicing	[n]	[p] or [t] or [k]	[b], [d], [g] as baselines	[a'kindinon ka'zino] 'harmless casinos'

The position of the stress and the number of the syllables in W1 and W2 were kept constant. The materials were embedded in five conditions, which manipulated the strength of the prosodic boundary between W1 and W2 by using different syntactic constructions and length manipulations for the utterances involved. These were intended to elicit gradually stronger prosodic levels (e.g. W1 and W2 belonging to the same or different Adj Phrases within the same clause, or to different Adj Phrases across short and long clauses, etc.). A total of 2150 tokens were analysed, approximately 1300 for vowel hiatus, and 850 for stop voicing.

For vowel hiatus, duration measurements were taken of the vowels in hiatus (V1V2) as well as formant values ¼ and ¾ into the formant transitions. The results showed that the effect of boundary strength was significant: for the condition where the length of the utterances was manipulated V1V2 durations were longer than the rest of the conditions, and formant values resembled more those of the baselines. In order to quantify the percentage of occurrences of deletions/assimilations within each prosodic domain, and thus judge the effect of prosodic strength on the sandhi application, a perceptual experiment was designed, where

Greek listeners were asked to classify [ao] sequences as [ao], [aa] or [oo], and [oa] sequences as [oa], [aa], or [oo]. It was found that the exact percentage of instances which were classified as deletions/assimilations decreased gradiently across conditions, signalling a gradient effect of prosodic boundary strength (these results were corroborated by discriminant analyses). Interestingly, within each prosodic condition, instances of both V1 and V2 deletions occurred (contra Kaisse 1985), with V2 deletion being more common.

For stop voicing, the duration of the pre-boundary nasal, of the closure, and of the burst of the stop were measured. Also, amplitude measurements were taken at the mid point of the stops' closure and burst. The effect of boundary strength was significant for the duration of the nasal and for the closure of the stops, as well as for the amplitude of the closure and of the burst. A perceptual experiment was again performed in order to quantify the percentage of voiced occurrences within each prosodic domain. Stop voicing was relatively rare (1-9% occurrences were classified as voiced across domains), indicating that the effect was blocked in all conditions. Given that so few instances of stop voicing were identified, another domain was later added to the design (Article # Noun). In this domain, 58% of the instances were classified as voiced. This suggests that the rule is optional within this added domain, while the effect of prosodic boundary across domains is categorical, i.e. restricted to only one domain.

Overall, the effect of the prosodic boundary was similar on the raw acoustic measurements for the two processes in the sense that more extreme values for most measurements were found in higher domains. However, the fact that full voicing was blocked in higher domains while vowel deletions/assimilations were allowed across conditions indicates that the boundary effect was manifested differently across processes.

Taken together, these results suggest that the influence of the prosodic boundary on absolute phonetic values (such as duration and amplitude values) is gradient, and that, particularly for vowel hiatus, it is also gradient on the proportional distribution of resolution strategies across prosodic conditions. Vowel hiatus within conditions was resolved both non-categorically (in the way that was reported by Zsiga 1997), but also categorically, like traditional phonological analyses would predict. On the other hand, although stop voicing was similar to vowel hiatus in applying optionally within domains, it was categorically blocked across domains. Thus, the two processes were found to relate to the prosodic structure differently, with vowel hiatus partially reflecting a hierarchical structure (with the length of the utterances being also a significant factor), and stop voicing being present in a single domain (for similar findings see Tserdanelis 2005). The fact that the prosodic boundary effect appears to be gradient for some phenomena and not others elaborates on analyses as the one proposed by Nespor & Vogel (1986), and suggests that a gradience analysis, which takes a probabilistic approach to acoustic variation within and across prosodic domains, needs to be incorporated in order to capture differences in prosodic boundary effects between phenomena.

Acknowledgments

I am grateful to my supervisors, Dr. A. Turk and Prof. D. R. Ladd, for their useful comments and feedback. This work is supported by an ESRC Doctoral award (PTA-030-2005-00147).

References

Baltazani, M. (2006), Focusing, prosodic phrasing, and hiatus resolution in Greek. In L. Goldstein, D. Walen & C. Best (Eds.), *Proceedings of Laboratory Phonology VII* pp.473-494). Berlin: Mouton de Gruyter.

Casali, F. (1997). Vowel elision in hiatus contexts: Which vowel goes? *Language* 73(3), 493–533.

Fougeron, C. & Keating, P. (1997). Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America,* 101, 3728-3740.

Kaisse, E. M. (1985). *Connected Speech: The Interaction of Syntax and Phonology*. New York: Academic Press.

Nespor, M. & Vogel, I. (1986). *Prosodic Phonology*. Dordrecht: Foris Publications.

Tserdanelis, G. (2005). *The role of segmental sandhi in the parsing of speech: Evidence from Greek*. Doctoral dissertation Ohio State University.

Wightman, C., Shattuck-Hufnagel, S., Ostendorf, M. & Price, P. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America*, 91, 1707-17.

Zsiga, E. C. (1997). Features, gestures and Igbo vowel assimilation: an approach to the phonology/phonetics mapping. *Language*, 73, 227–274.

Social Selectivity in Adults' Novel Sound Learning

Eden Kaiser* & Benjamin Munson#

*Linguistics Program, University of Minnesota; kaise113@umn.edu
#Department of Speech-Language-Hearing Sciences, University of Minnesota; munso005@umn.edu

There is a good deal of evidence for the existence of social identity-based differences in speech (Labov 1972, and others), especially differences based on talker gender (Perry, Ohde, & Ashmead, 2001; Patterson & Werker, 2002; Foulkes, Docherty, & Watt 2005, and others). However, the nature of the process by which these differences are acquired is still unknown. The purpose of this study is to investigate sex-specific phonetic learning in a laboratory-based novel-sound learning task. Specifically, we examined male and female participants' tendencies to emulate phonetically distinct variants of novel sounds produced by talkers who shared their biological sex, as opposed to the variant produced by the opposite sex. We believe that the outcome of this endeavor will be a better understanding of how these variables might be acquired during first-language acquisition.

To examine this, we created one set of one-syllable nonwords whose onsets contained a voiceless lateral obstruent, and another set whose onsets contained a velar fricative. These non-English sounds were used in order to control for any previous sex-related biases people may have for familiar English sounds. Audio productions of these novel words were recorded by two trained linguists whose voices the authors judged as being prototypically female and male. The novel words were presented one at a time, by either the female voice or the male voice, and were associated with novel objects shown on a computer screen. Each participant was trained on one onset which the man and woman produced differently (hereafter called "sex-biased"), and one onset which they produced equally variably (hereafter called "random"). The two variants for the voiceless lateral obstruent were the voiceless lateral fricative [ɬ] and the voiceless lateral affricate [tɬ]. The two variants for the velar fricative were the voiced velar fricative [ɣ] and the voiceless velar fricative [x]. Participants were systematically exposed to variations of onsets: half the participants heard sex-biased variation for the lateral onset and random variation for the velar onset, while the other half heard random variation for the lateral onset and sex-biased variation for the velar onset. Three iterations of learning phases followed by test phases were presented to each participant, first for one onset, then for the other. During the learning phases, participants saw a novel object on the screen and heard over headphones "This object is a ____", spoken by either the man or the woman, and onsets varied accordingly. During the test phases, participants were asked to name objects whose pictures appeared on the screen one at a time. Below is the counter-balancing scheme used to systematically expose participants to both sex-biased and random variation. There were five women and five men in each of four groups.

Group 1	random [ɬ]/[tɬ]	female-biased [ɣ] / male-biased [x]
Group 2	random [ɬ]/[tɬ]	female-biased [x] / male-biased [ɣ]
Group 3	female-biased [tɬ] / male-biased [ɬ]	random [x]/[ɣ]
Group 4	female-biased [ɬ] / male-biased [tɬ]	random [x]/[ɣ]

Table 1: Sex-biasing counter-balancing scheme

Results were coded in narrow phonetic transcription by an experienced transcriber who was blind to the purposes of the experiment. The dependent measures were calculated by one of the authors using the following criteria: onsets which contained a fricative were judged to be approximating the lateral fricative [ɬ], whereas onsets which contained an obstruent were judged to be approximating the lateral affricate [tɬ];

and onsets which contained a voiceless velar or voiceless glottal were judged to be approximating the voiceless velar fricative [x], whereas all onsets which contained either a voiced velar/glottal or a rhotic were judged to be approximating the voiced velar fricative [ɣ].

The authors found that, overall, both men and women tended to produce more approximations of the onsets [tɬ] and [x] over their respective variants, [ɬ] and [ɣ].

<table>
<tr><th></th><th>Females' use of [tɬ]</th><th>Males' use of [tɬ]</th><th>Females' use of [ɣ]</th><th>Males' use of [ɣ]</th></tr>
<tr><td>Group 1</td><td rowspan="2">71.212</td><td rowspan="2">79.059</td><td>18.519</td><td>6.944</td></tr>
<tr><td>Group 2</td><td>14.293</td><td>4.646</td></tr>
<tr><td>Group 3</td><td>64.583</td><td>67.677</td><td rowspan="2">3.819</td><td rowspan="2">4.444</td></tr>
<tr><td>Group 4</td><td>59.722</td><td>44.556</td></tr>
</table>

Table 2: Results (in percentages)

The influence group and training condition on the percentage use of different variants did not achieve statistical significance using conventional alpha levels. However, there were some noteworthy trends in the data. For the lateral onset [ɬ]/[tɬ], women and men were more likely to attempt to produce [ɬ] when presented with sex-biased variation (either biased for *or* against production of [ɬ]). More interestingly, men were much more likely to produce [tɬ] than [ɬ] when presented with random variation, than when biased *against* producing [tɬ]. For the velar onset [x]/[ɣ], males rarely attempted to approximate the voiced velar fricative, opting instead for the unvoiced variant. Females were about four times as likely to attempt to approximate the voiced velar fricative when they were presented with sex-biased variation (especially when biased for the voiced variant). But when females were presented with random variation, their attempts to approximate the voiced variant were on par with the men's.

Based on these results, we can conclude that women and men are sensitive to sex-biased differences when acquiring novel sounds. In other words, people learn new phonetic variants selectively, based on the social-indexical characteristics associated with the phonetic input, namely, one's own sex and the sex of the speaker. The next step along this line of research is to reproduce these results, removing the potentially complicating factor of production. Currently, we are designing a perception experiment to examine whether adults' perception is similarly biased by the perceived sex of the adults who model the novel sound. The result of these two experiments together will be an improved understanding of the extent to which adults' sound learning is indexed to perceived social characteristics of the people producing the sounds being learned.

References

Foulkes, P., Docherty, G., & Watt, D. (2005). Phonological variation in child directed speech. *Language*, 81, 177-206.

Labov, W. (1972). *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

Patterson, M. L. and Werker, J. F. (2002). Infants ability to match dynamic phonetic and gender information in the face and voice. *Journal of Experimental Child Psychology*, 81, 93-115.

Perry, T., Ohde, R., and Ashmead, D. (2001). The acoustic bases for gender identification from children's voices. *JASA*, 109, 2988-2998.

The role of phonetic detail in associating phonological units

Vsevolod Kapatsinski* and David B. Pisoni#

*Department of Linguistics, Cognitive Science Program, Indiana University, USA; vkapatsi@indiana.edu
#Department of Psychological and Brain Sciences, Indiana University, USA; pisoni@indiana.edu

A large part of phonological competence is usually assumed to consist of the knowledge of co-occurrence relations between phonological units. Native speakers know which phoneme combinations and phoneme-morpheme sequences are legal in their language, and which are not, as well as the frequencies with which particular sequences occur (e.g., Pierrehumbert 1994). A speaker can be said to possess knowledge of the co-occurrence relation that holds between two phonological units if s/he can generalize to instances of the relevant phonological units in unfamiliar contexts. For instance, if an English speaker possesses a rule that prohibits labial codas following /aʊ/ (but not any other vowel), s/he should judge monosyllabic words with labial codas to be unacceptable iff they contain the phoneme /aʊ/. Thus, to be said to know the rule, the subject must extract [aʊ] from the novel acoustic signal and recognize it an instance of /aʊ/.

In the present paper, we show that the likelihood with which a newly-learned rule involving a particular phoneme is extended to a novel instance of the phoneme depends on how easy it is to classify the novel instance as belonging to the phoneme in question, indicating that the process of even unconscious phonemic categorization is not completely automatic. Yet, the rule is applied to novel syllables as accurately as to syllables presented during training, indicating that subjects do form a generalization at the phonemic level as opposed to simply memorizing exemplars of whole stimuli presented during training. The results suggest that learners generalize to the phonemic level of abstraction but the process of phonemic categorization is not error-free, hence the likelihood that a phoneme's associations will be accessed (and strengthened) depends in part on whether the phoneme's identity will be detected in time.

In the first study, native English speakers were exposed to an artificial language in which CVC stems sharing a particular rime (VC) or body (CV) also shared an affix. One group of participants learned that /Cæʃ/ stems take the affix /mɪn/ while /Cʌg/ stems take the affix /num/. Thus, participants assigned to this group were exposed to rime-affix co-occurrences. The other group was exposed to body-affix co-occurrences where /ʃæC/ stems took /mɪn/ while /gʌC/ stems took /num/. Knowledge of the co-occurrence relations was tested by presenting the participants with novel syllables containing familiar rimes or bodies and asking for the affix. Whether the affix came after or before the stem had no significant effect. However, as Figure 1 shows, rime-affix associations were much easier to form than body-affix associations (by subjects: t=5.401, df=66, p<.0001, by items: t=13.445, df=42, p<.0001). The subjects' accuracy with novel syllables was as high as with familiar syllables (F(1,66)=.002, p=.98), indicating that the subjects did not simply memorize the tokens presented to them but rather formed rime-affix associations. However, there was significant within-category variability: rule application was more accurate with some novel stems containing a given familiar string than with other novel stems containing the same familiar string.

To explain this between-item variability, we presented a new group of native English speakers with the stem syllables used in the first study and asked them to decide whether the vowel they are hearing is /ʌ/ or /æ. Examples of minimal pairs containing /ʌ/ and /æ/ were provided (e.g., /kʌt/-/kæt/). Like in the original study, syllables were presented over headphones in the absence of noise. While accuracy in the task was at ceiling, syllables differed in how fast the identity of the syllable's vowel was detected. We then examined vowel categorization reaction times in response to syllables that served as generalization stimuli in the first study and correlated reaction time in vowel categorization with generalization accuracy. As Figure 1 shows, there was a strong linear correlation. Generalization accuracy is low when the generalization stimulus is difficult to categorize into the same phonemic category as the training stimuli.

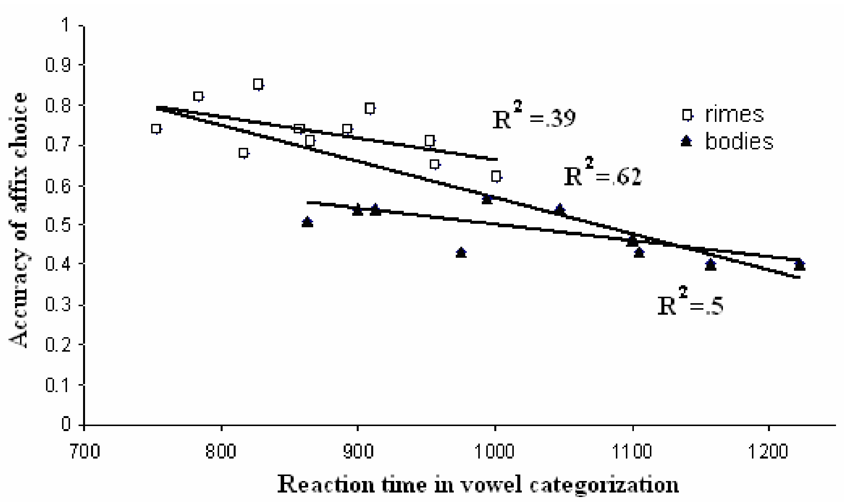


Figure 1: Reaction time in vowel categorization correlates with accuracy in generalizing CV-affix or VC-affix associations to the syllable.

This finding provides a potential account for recent findings that associations between vowels are more difficult to learn than associations between consonants (Bonatti et al. 2005, Creel et al. 2006). Since instances of a vowel phoneme are both acoustically and perceptually more variable than instances of a consonant phoneme, generalization to a novel instance of a vowel phoneme is predicted to be more difficult than generalization to a novel instance of a consonant phoneme. The same explanation could also be offered for the difference in associability between bodies and rimes. If listeners do not categorize incoming speech into phonemic categories automatically, and variations in the coda have a greater impact on vowel quality than variations in the onset, the equivalence of different tokens of the same rime may be easier to detect than the equivalence of different tokens of the same body, making rime-affix associations easier to acquire and generalize than body-affix associations. However, Figure 1 shows that variation in phonemic categorization does not account for all the variation in associability that the rime/body distinction accounts for (R^2=62% vs. 50+39=89%). We propose that a phoneme string A is easy to associate with something else when its instances are perceptually similar and when an instance of A present in the acoustic signal is likely to be extracted from the signal due to instantiating a constituent in the learner's language. Together these two factors account for most of the between-item variability in the present study (89%).

References:

Bonatti, L., Peña, M., Nespor, M., & Mehler, J. (2005). Linguistic constraints on statistical computations: The role of consonants and vowels in continuous speech processing. *Psychological Science*, 16, 451-9.

Creel, S.C., Aslin, R.N., & Tanenhaus, M.K. (2006). Acquiring an artificial lexicon: Segment type and order information in early lexical entries. *Journal of Memory and Language*, 54, 1-19.

Pierrehumbert, J. (1994). Syllable structure and word structure: A study of triconsonantal clusters in English. In P. Keating (Ed.), *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology III* (pp.168-190). Cambridge: Cambridge University Press.

Rule Reliability and Productivity: Velar Palatalization in Russian and Artificial Grammar

Vsevolod Kapatsinski

Department of Linguistics and Cognitive Science Program, Indiana University, USA
vkapatsi@indiana.edu

Russian velar palatalization (k→tʃ, g→ʒ, x→ʃ) presents an intriguing case of a morphophonological alternation that is exceptionless for a given suffix, yet not fully productive in natural loanword adaptation with some of the suffixes. In particular, while the process is shown to be only partially productive before –i and –ik but fully productive before –ok and –ek, contrary to the naturalness of the suffixes as triggers of palatalization. A model of rule induction and weighting (the Rule-based Learner, developed by Albright and Hayes 2003) is trained on the lexicon of Russian verbs, in which velar palatalization is exceptionless (always applying before the stem extension –i), and tested on new borrowings and nonce probes from elicited production tasks. Despite the fact that velar palatalization is exceptionless in the training set, it is correctly predicted to often fail with novel words based on information in the lexicon.

When a foreign verb is borrowed into Russian, it must be assigned a stem extension, e.g., upload → /ʌplod+i+t^j/, duck → /dʌk+a+t^j/, lock → /lotʃ+i+t^j/. The two most productive stem extensions in modern Russian are –i and –a. While –i always triggers velar palatalization in the native lexicon, -a does not, e.g., duck →/dʌk+a+t^j/, lock → /lotʃ+i+t^j/. Importantly, velar-final roots favor –a while –i is favored elsewhere.

When exposed to the native lexicon, the Rule-based Learner discovers the following crucial rules: C→C+a, C→C+i, k→tʃ+i, g→ʒ+ɨ. In this model, the likelihood that a rule will apply is determined by its reliability relative to competing rules. Reliability is defined as the number of words to which the rule applies divided by the number of words to which it could apply. For instance, the reliability of the rule k→tʃ+i is the number of k-final roots that take –i and change the final consonant to the total number of k-final roots in the lexicon. Since velars favor –a over –i while most consonants favor –i over -a, the rules k→tʃ+i and g→ʒ+ɨ, which involve velar palatalization, are not very reliable relative to the rule C→C+i, which leaves the consonant unchanged. Thus, velar palatalization is correctly predicted to fail before -i. On the other hand, the diminutive suffixes –ek and –ok usually attach to velar-final inputs, hence rules stipulating velar palatalization before these suffixes are much more reliable than the rules that stipulate that the suffix can be attached with no change to the preceding consonant.

These predictions hold regardless of whether the stem change and the suffix are chosen during a single decision stage or the choice of the suffix precedes the decision on whether to change the stem. However, only the former (single-stage) model is able to account for segmental context effects on palatalization. As Figure 1 shows, the single-stage model successfully predicts that velar palatalization should fail more often when the stem ends in a consonant cluster than when it ends in a VC sequence, and more if it contains a front vowel than if it contains a back one. Importantly, these predictions are made only if speakers choose the suffix and the stem change at the same time. If the decision to apply the stem change is made after the affix is chosen (Stage I: choose –i, Stage II: choose whether to palatalize), the model predicts no influence of penultimate segment identity. Thus, I argue that the suffix should not be seen as a 'trigger' of velar palatalization but rather that both the suffix and the stem change are markers of the function typically associated with the suffix and are chosen simultaneously.

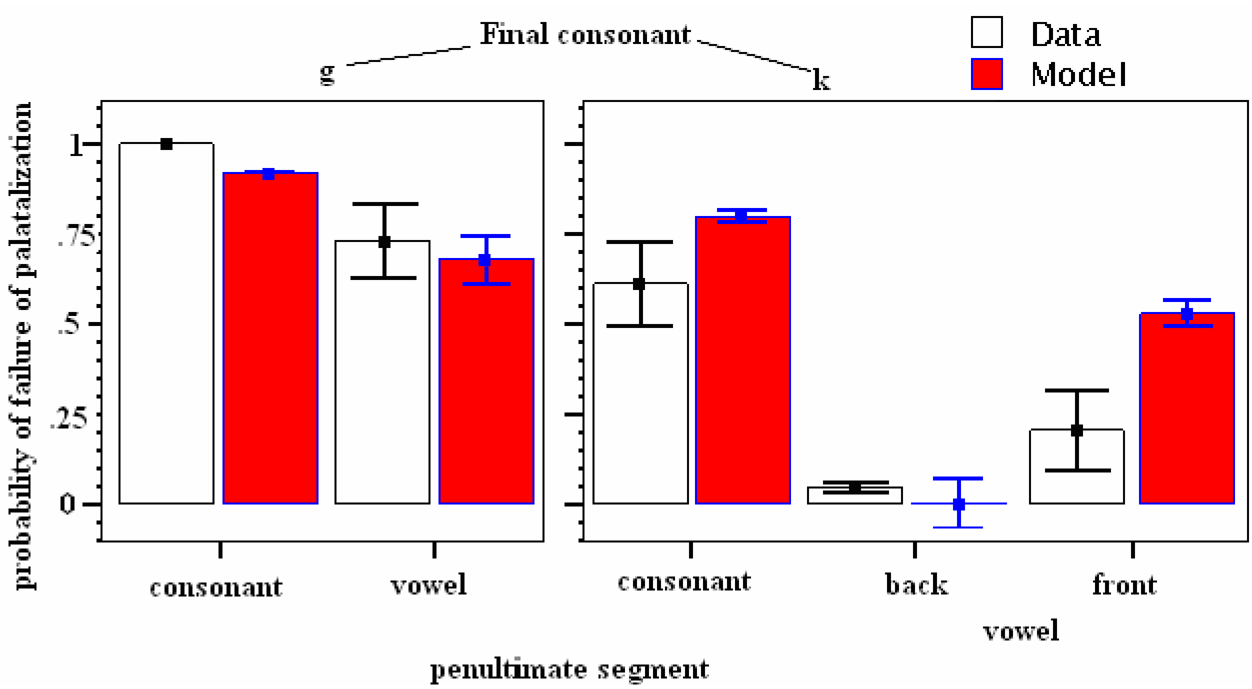


Figure 1: Observed and predicted rates of velar retention as a function of final consonant and penultimate segment type (predicted rates rescaled to have the same variance as the data).

The importance of rule reliability for productivity is confirmed by an artificial grammar learning experiment, in which native English speakers were exposed to one of two languages. Both languages contained two plural suffixes, -i and –a. In both languages, 30 words exhibited velar palatalization, which was exceptionless before –i and never occurred before –a. The same total number of training trials was presented. However, in Language I, -i occurred mostly with velar-final inputs whereas in Language II non-velar-final inputs also took –i most of the time. Thus, the languages differed only in the reliability of the no-change rule C→Ci, which was higher in Language II than in Language I. As expected, velar palatalization was much more likely to fail before –i in Language II than in Language I (63% vs. 29% respectively).

Rule reliability does not provide a complete account of productivity. Thus, in the Russian loanword adaptation data, it was observed that velar palatalization before –i was blocked if its application would produce a homonym with an existing Russian verb. This base recoverability effect is not predicted by the Rule-Based Learner. In the artificial-grammar learning experiment, subjects exposed to either language learned that singulars ending in /tʃ/ or /dʒ/ take –i, despite having seen no examples in which this happens (singulars ending in /tʃ/ or /dʒ/ were not presented during training). Evidently, subjects generalized over the plurals, which featured /tʃi/ and /dʒi/ but not /tʃa/ and /dʒa/. Since the Rule-Based Learner can only generalize over input-output pairings, it does not predict this product-oriented effect. Thus, I conclude that rule reliability is an important influence on productivity but not the only one, interacting with base recoverability/homonymy avoidance and reliability of product-oriented generalizations.

References

Albright, A., & Hayes, B. (2003). Rules vs. analogy in English past tenses: A computational / experimental study. *Cognition*, 90, 119-61.

Compensation ?=? Mental Representation

Shira Katseff* & John Houde#

*Department of Linguistics, University of California at Berkeley, USA; skatseff@berkeley.edu
#Department of Otolaryngology, University of California at San Francisco, USA; houde@phy.ucsf.edu

The speech accommodation literature demonstrates that listening and speaking are intimately connected: what a talker hears from an interlocutor will affect what that talker says (Pardo, 2006). The same turns out to be true of self-produced speech: hearing an altered version of one's own voice changes speech production. This phenomenon, speech adaptation, is used to investigate the role of phonetic information in lexical planning.

Speech adaptation is easiest to understand by analogy to a parallel sensorimotor adaptation experiment. First, a subject reaches for an object sitting on a table. Then that subject puts on a pair of prism glasses which shift the visual field, causing the object to appear to be in a different position. At first the subject fails to grasp the object, but soon adjusts his/her target and is able to reach the object in spite of altered visual feedback. This is compensation. Just after removing the glasses, the subject continues to expect the shifted visual field, and reaches past the object several times before being able to grasp it again.

In speech adaptation, subjects wear a headset. They speak into the microphone and hear their speech played back to them through the earphones. The auditory version of the task used here involves four stages: baseline, ramp, plateau, and adaptation. Subjects repeat a single word, in this case, ‘head', over a large number of trials. In “reaching” sensorimotor adaptation experiments, subjects initially see the object on the table in its true position. In speech adaptation experiments, subjects hear their voices unaltered during the baseline stage. During each trial in the ramp stage, auditory feedback is altered a small amount until it reaches a maximum value. Feedback alteration is held at that maximum value during the plateau stage. In this experiment, there are 5 sets of ramps and plateaus, after which feedback drops suddenly back to normal for the adaptation stage.

Previous work shows that subjects generally change their speech to oppose the auditory feedback change. For example, when F1 in auditory feedback is raised, making their /ɛ/ sound more like an /a/, subjects compensate by speaking with a lower F1; the vowels they produce sound more like /ɪ/. Similar experiments show that subjects will compensate for alterations in F0, F1, and F2 feedback, indicating that all three of these formants are important to a speaker's representation of the target utterance (e.g., Burnett, Freedland, Larson & Hain, 1998; Houde & Jordan, 2002, Purcell & Munhall, 2006).

This paper investigates the relative importance of these three formants to the representation of ‘head'. Subjects participated in speech adaptation experiments for F0, F1, and F2 shifts on three different days. Formants were altered in randomized order, and all data was analyzed relative to a control condition where the subject went through the experiment without auditory feedback alteration.

Several trends in the data are clear. First, compensation is never complete. For example, in response to a total feedback increase of 100 Hz in F1, a subject might lower his or her F1 by 50 Hz. Second, compensation is more complete for small feedback shifts than for large feedback shifts. Third, compensation is more complete for F1 shifts than for F2 shifts.

These results suggest that both acoustic and sensorimotor feedback are part of one's lexical expectation. Because auditory feedback is altered while motor feedback is not, feedback from these two sources can conflict. For small shifts in auditory feedback, the amount of potential conflict is small and the normal motor feedback does not affect compensation. But for large shifts in auditory feedback, the amount of conflict is large. Abnormal acoustic feedback pushes the articulatory system to compensate, and normal motor feedback pushes the articulatory system to remain in its current configuration, damping the compensatory response.

The way in which subjects compensate suggests that the lexical target vowel obligatorily involves multiple formants. In response to a change in F1, many subjects change their production of both F1 and F2:

when their F1 feedback is increased, they produce a lower F1 and a higher F2. The exact proportion of compensatory F1 and F2 produced is not constant and depends on the amount of feedback shift. The converse is true of F2 feedback alteration. When F2 feedback is raised, subjects compensate by lowering their F2 and raising their F1. However, compensation for F0 shifts is orthogonal to compensation for the other two formants: shifting F1 and F2 feedback had no effect on F0 production, and compensation for shifts in F0 feedback were not systematically related to changes in F1 or F2 production.

One possible explanation of these results is a joint representation of a vowel's F1 and F2 in either perception or production. Ongoing experimental work is investigating whether the joint movement of these two formants can be attributed to a combination of acoustic and motor constraints.

To determine whether changes in expectation for the vowel in ‘head' affects vowels in phonologically similar words, subjects produced a set of /hVd/ minimal pairs before and after each experiment. Several subjects changed their production of the adjacent minimal pairs ‘had', ‘hod', and ‘hoed'. Thus speech adaptation causes generalization not only to future productions of the same word, but also to adjacent words. These results are analogous to lexical priming, and suggest that the internal model of acoustic speech expectations is bound up with linguistic information.

Acknowledgements

We would like to thank Keith Johnson for invaluable advice and assistance with data measurement. Many thanks also to Richard Hahn, Reiko Kataoka, Grant McGuire, and the UC Berkeley phonology lab for their enthusiasm and general wisdom.

References

Burnett, T. A., Freedland, M. B., Larson, C. R., & Hain, T. C. (1998). Voice F0 responses to manipulations in pitch feedback. *Journal of the Acoustical Society of America*, 103 (6), 3153-3161.

Houde, J. F., & Jordan, M. I. (2002). Sensorimotor adaptation of speech I: Compensation and adaptation. *Journal of Speech, Language, and Hearing Research*, 45 (2), 295-310.

Jones, J. A., & Munhall, K. G. (2000). Perceptual calibration of f0 production: Evidence from feedback perturbation. *Journal of the Acoustical Society of America*, 108 (3), 1246-1251.

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America*, 119(4), 2382-2393.

Purcell, D. W., & Munhall, K. G. (2006). Compensation following real-time manipulation of formants in isolated vowels. *Journal of the Acoustical Society of America*, 119 (4), 2288-2297.

Interactive and autonomous modes of speech perception: Phonological knowledge and discrimination in English and French listeners

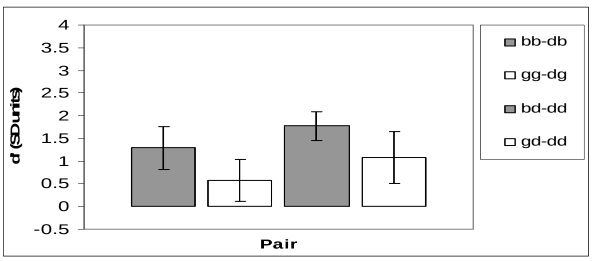
Michael Key

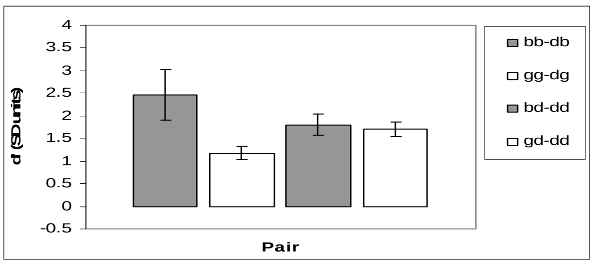
Department of Linguistics, University of Massachusetts, Amherst

The processing of the incoming speech signal can occur in two distinct general *modes*: the *autonomous mode* and the *interactive mode*. The autonomous mode of processing is characterized by an apparent lack of influence from sources of linguistic knowledge or bias (e.g. the lexicon) in certain perceptual tasks in which an influence is expected. For example, recent data from Kingston, Mash, Chambless, Kawahara, Katz, & Key (submitted) shows that wordhood biases do *not* make pairs drawn from a word↔non-word continuum more cumulatively discriminable than pairs drawn from a non-word↔non-word continuum. In contrast, the interactive mode is characterized by the influence of linguistic knowledge (i.e. feedback) in a perceptual task in which an influence is expected. A classic example of the interactive mode are the lexical biases on phoneme identification reported by Ganong (1980). While it is true that the literature is full of studies that report interactive processing (i.e. some linguistic bias) and relatively devoid of studies reporting autonomous processing (i.e. the lack of a bias where one would be expected), it is also the case that the possible combinations of *task type* and *source of linguistic influence* has not nearly been exhausted. In particular, the large majority of these studies have relied on *recognition* tasks (e.g. phoneme identification, word detection, etc.) and have tested for various influences directly from the lexicon (e.g. wordhood, lexical statistics). As a result, we do not yet have a good understanding of what determines which mode of processing is observed.

The study reported here fills in one of these missing combinations by using *discrimination* tasks to test for the influence of phonological knowledge using non-word stimuli, thus factoring out possible lexical influences. English has a well-known pattern of regressive place assimilation of coronal stops to labials and velars when the following context is labial or velar (e.g. Wells, 1982): e.g. *good girl → goo***[g]** *girl*. Previous studies using recognition tasks have shown that English listeners recover the identity of a phoneme that matches the lexical representation (Gaskell & Marslen-Wilson, 1996), with a replication in non-word stimuli (Gaskell & Marslen-Wilson, 1998). This study tested the hypothesis that discrimination tasks would contrast with the recognition results in producing an autonomous mode of processing. Specifically, English listeners should have no more difficulty discriminating coronal from non-coronal stops more when the following context is *assimilation-viable* (i.e. non-coronal), then when it is unviable (i.e. coronal). In this experiment, the voiced oral stops [b, d, g] were presented in the context preceding heterosyllabic [b, d, g] in non-word stimuli of shape VC.CV. On each trial, listeners' task was to decide whether the pair of stimuli heard were the same or different; the format of discrimination was AX ('same-different').

Figure 1 shows mean discrimination performance when C_1 varies between a coronal and a non-coronal. The left pair of bars represent the pairs in which both a coronal was present and the following stop context was viable ([b] or [g]). Discrimination of [b] and [d] in the viable labial assimilation context is compared with discrimination of [b] and [d] in the unviable coronal context (gray bars), and discrimination of [g] and [d] in the viable velar context is compared with that in the unviable coronal context (white bars). Discrimination in both viable contexts is significantly poorer, as confirmed by paired one-tailed t-tests ($t(15) = 2.73$, $p = 0.016$ in the labial context, $t(15) = 2.28$, $p = 0.037$ in the velar context).





Figures 1 (left) and 2 (right): Compared d' values in English listeners (Fig. 1) and French listeners (Fig.2) between pairs in which C2 is a viable assimilation context and pairs in which C2 is not a viable assimilation context (95% CIs).

In contrast, French lacks this pattern of place assimilation, and thus no difference in discriminability is predicted for French listeners (Figure 2). Darcy, Ramus, Christophe, Kinzler, & Dupoux (to appear) found that English listeners compensated for place assimilation significantly more often than French listeners in a phoneme monitoring task. In contrast, French listeners perform better in the viable labial context than in the unviable coronal context ($t(7) = 3.01$, $p = 0.019$), but perform worse in the viable velar context ($t(7) = 7.45$, $p < 0.001$). The difference between the performance of English and French listeners in the viable labial context is most telling for the predictions of our hypothesis because it shows that AX discrimination apparently induces an interactive mode of processing whose source is language-particular phonological knowledge.

In a follow-up of this experiment, the stimuli remained the same but the format was changed to 4IAX, which has previously been shown to encourage an 'auditory' (i.e. non-phonologically-biased) mode of processing (Gerrits & Schouten, 2004). Figure 3 compares discrimination of coronals from non-coronals in both viable (non-coronal) and unviable (coronal) contexts by English listeners.

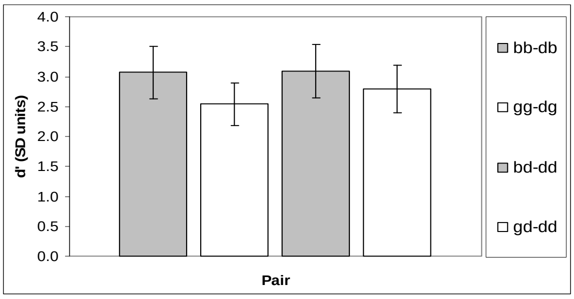


Figure 3: Mean discrimination of coronals from non-coronals in viable and non-viable contexts (95% CIs).

Discrimination is no worse in viable contexts than in unviable contexts and discrimination of coronals from non-coronals is no worse than discrimination of two non-coronals. discrimination of [b] from [d] is no worse in a viable context (before [b]) than in an unviable context (before [d]) ($t(15) = 0.78$, $p > 0.10$); discrimination of [b] from [d] is no worse than [b] from [g] with viable context held constant ($t(15) = 1.27$, $p > 0.10$). Thus, the results from the 4IAX discrimination task show an autonomous mode of processing.

The autonomy-interaction debate has often been cast in terms of *models* of speech perception rather than *modes*, as in this paper. There are two reasons I avoid the debate over models. First, arguments for one type of model over another are often fueled by the claim that effects are ‘early’ or ‘late’. Although I have not done a comprehensive survey of these arguments, it is far from clear what point in the course of processing divides ‘early’ from ‘late’; these are terms of art rather than well-defined stages in a model. Second, the choice of model is ultimately a question of parsimony; autonomous models are often claimed to be more complex because of the additional level of processing they posit. However, autonomous and interactive models should not be compared in this way unless specific implementations (i.e. mechanisms) of each are discussed. In short, invoking Ockham’s Razor is appropriate when choosing between fully explicit models, but inappropriate when choosing between *classes* of models.

Selected References

Darcy, I., Ramus, F., Christophe, A., Kinzler, K., & E. Dupoux (to appear). Phonological knowledge in compensation for native and non-native assimilation In: F. Kügler, C. Féry & R. van de Vijver (eds.) *Variation and Gradience in Phonetics and Phonology*. Berlin: Mouton De Gruyter.

Gaskell, M.G., & Marslen-Wilson, W.D. (1996). Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception and Performance,* 22, 144-158.

Gaskell, M.G., & Marslen-Wilson, W.D. (1998). Mechanisms of phonological inference in speech perception. *Journal of Experimental Psychology: Human Perception and Performance,* 24, 380-396.

Gerrits, E., & Schouten M.E.H. (2004). Categorical perception depends on the discrimination task. *Perception and Psychophysics,* 6, 363-378.

Kingston, J., Mash, D., Chambless, D., Kawahara, S., Katz, J., & M. Key (submitted). Lexical and spectral contextual effects and auditory autonomy in the perception of neighboring vowels and consonants. *Language and Speech*.

Phonetic Cues to Gemination in Lebanese Arabic

Ghada Khattab & Jalal Al-Tamimi

Speech and Language Sciences Section, Newcastle University, UK;
ghada.khattab@ncl.ac.uk; jalal.al-tamimi@ncl.ac.uk

This paper reports on phonetic and phonological patterns of gemination and vowel length in Lebanese Arabic (LA). Both short and long vowels occur before singleton and geminate consonants in LA leading to the following permissible syllable structures:

/ˈtˤabaʔ/ ‘*dish*’	/ˈtˤaabaʔ/ ‘*he matched*’	/ˈtˤabbaʔ/ ‘*he persuaded*’	/ˈtˤaabbe/ ‘*she is bending over*’

The phonetic realisation of the geminate contrast and the temporal relationship between medial consonants and their surrounding vowels has been the subject of many cross-linguistic and cross-dialectal studies (e.g. Al-Tamimi, 2004; Arvaniti, 2001; Ghalib, 1984; Ham, 2001; Hassan, 2003). While the emphasis has mainly been on durational cues to gemination, some studies have suggested that other non-temporal characteristics contribute to the perceptual effect of gemination. These include a palatalised resonance for geminate sonorants (Local & Simpson, 1988) and palatal contact for geminate stops (Payne, 2005), laminal contact for geminates as opposed to apical contact for singletons (Payne, 2006), a flatter shape of the tongue in geminate articulation (Payne, 2006), more lenited stops in singleton contexts (Ridouane, 2007; Ladd & Scobbie, 2003), and lower burst amplitude and occasional absence of bursts in singleton stops (Local & Simpson, 1999; Ridouane, 2007). Some of these cues have led researchers to suggest a tense/lax distinction between singleton and geminate consonants alongside the durational contrast; this is thought to enhance the perceptual distance between singletons and geminates. The domain of gemination has also been found to extend to the surrounding vowels and sometimes across the whole word. For instance, preceding vowels have been found to be longer and more centralized before singleton than before geminate consonants (Local & Simpson, 1988), and whole words have been found to be produced with an overall lax setting in singleton as opposed to geminate contexts.

This study contributes to the literature on gemination by providing a detailed examination of LA. There are very few phonetic studies of LA (Nasr, 1960), and only a pilot study on the acoustic patterns of consonant and vowel length in the colloquial variety (Khattab, 2007). While consonant gemination in LA is very frequent (all 27 LA consonants can be geminates) and plays an important morpho-syntactic role in the language, little is known about the phonetic realisation of singleton and geminate targets in this dialect or about the role played by the preceding vowel. The same is true regarding phonemic vowel length. And while most studies on gemination in Arabic have concentrated in durational cues to the singleton/geminate contrast, this study looks at a variety of non-temporal cues in order to investigate whether these play an important role in the implementation of this phonological contrast.

Twenty Lebanese males and females were recorded reading target word-lists containing medial singleton and geminate consonants (nasals, fricative, liquids, and approximants) preceded by long and short vowels. Acoustic and auditory analyses of medial consonants (C(C)) and of preceding (V1) and following (V2) vowel durations were made. Temporal measurements included V1, V2, and medial C(C) duration. Non-temporal measurements included formant frequencies at mid-point and offset of V1, mid-point of sonorant consonants, onset and mid-point of V2, and intensity and $f0$ in V1, C(C) and V2.

Temporal results suggest a robust role for duration in distinguishing between short and long consonants and vowels in LA. There were separate durational distributions for singleton and geminate consonants and for target short and long vowels. The duration of geminate consonants in this study is generally comparable to what has been found for Jordanian (Al-Tamimi, 2004), Iraqi (Hassan, 2003), Berber (Ridouane, 2007), and Malayalam (Local & Simpson, 1988). However, it is much shorter than what has been found for other languages such as Greek (Arvaniti, 2001), Swedish (Hassan, 2003), and Finnish (Kunnari, Nakai, & Vihman, 2001). A surprising result for duration is that consonant and vowels did not exhibit temporal compensation at the absolute durational level, i.e. vowels were not shorter before geminate than singleton consonants as is often found in the literature. This might be due to the fact phonological length plays an important role in vowels as well as consonants in Arabic, but temporal compensation has been found to occur in Iraqi and Jordanian Arabic (e.g. Al-Tamimi, 2004; Hassan, 2003) so this matter will be investigated further.

Non-temporal results also show a surprising lack of difference in the spectral cues in vowels preceding and following singleton and geminate consonants, suggesting a lack of vowel quality difference in the implementation of this contrast. There is a slight tendency for vowels surrounding geminate consonants to be raised and fronted compared with vowels surrounding singleton consonants, but this is more prominent in the data for females than males. Moreover, vowel quality, intensity and *f*0 vary more depending on whether the preceding vowel is phonemically long or short rather than depending on the phonemic length of the consonant itself.

More work is underway to look at stops since some differences have been reported in VOT, burst amplitude and consonant and vowel intensity, but results so far highlight the role of temporal cues over spectral and other non-temporal cues such as tense/lax realisations in the distinction between singleton and geminate consonants in LA. This may suggest that the underlying contrast for gemination in LA is mainly temporal and its domain is restricted the 2nd syllable.

References

Al-Tamimi, F.Y. (2004). An experimental phonetic study of intervocalic singleton and geminate sonorants in Jordanian Arabic. *Al-Arabiyya*, 29, 37-52.

Arvaniti, A. (2001). Comparing the phonetics of single and geminate consonants in Cypriot and Standard Greek. *Proceedings of the 4th International Conference on Greek Linguistics* (pp. 37–44). Thessaloniki: University Press.

Ghalib, G. B. M. (1984). *An Experimental Study of Consonant Gemination in Iraqi Colloquial Arabic*. Unpublished Ph.D. Thesis, University of Leeds.

Ham, W. (2001). *Phonetic and phonological aspects of geminate timing*. New York: Routledge.

Hassan, Z. M. (2003). Temporal compensation between vowel and consonant in Swedish & Arabic in sequences of CV:C & CVC: and the word overall duration. *PHONUM*, 9, 45-48.

Khattab, G. (2007). A phonetic study of germination in Lebanese Arabic In Jürgen Trouvain & William J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences*. Saarbrücken, Germany.

Kunnari, S. Nakai, S. Vihman, M. M. (2001). Cross-linguistic evidence for the acquisition of geminates. *Psychology of Language and Communication,* 5, 13-24.

Local, J., & Simpson A. (1988). The domain of gemination in Malayalam. In D. Bradley, E.J.A. Henderson, & M. Mazaudon (Eds.), *Prosodic Analysis and Asian Linguistics: To Honour R. K. Sprigg* (pp. 33–42). Pacific Linguistics Series C - No. 104.

Nasr, R.T. (1960). Phonemic length in Lebanese Arabic. *Phonetica,* 5, 209-211.

Payne, E. (2006). Non-durational indices of Italian geminates: an EPG study. *Journal of the International Phonetic Association*, 36(1), 83-95.

Payne, E. (2005). Phonetic variation in Italian consonant germination. *Journal of the International Phonetic Association*, 35(2), 153-189.

Ridouane, R. (2007). Gemination in Tashlhiyt Berber: an acoustic and articulatory. *Journal of the International Phonetic Association,* 37(2), 119-142.

The independence of auditory and categorical effects on speech perception

John Kingston*, Shigeto Kawahara#, Della Chambless%, Michael Key*, and Sarah Watsky*

*Linguistics Department, University of Massachusetts, Amherst, jkingston@linguist.umass.edu;
#University of Georgia; %University of North Texas

Does the listener's linguistic knowledge feed back to all earlier stages of processing speech perception (McClelland & Elman, 1986) or is the initial processing of the speech signal uninfluenced by such knowledge (Norris, et al., 2000)? An autonomous linguistically naïve stage deals in a different currency, continuous auditory qualities, from the categories that linguistically informed processing deals in.

The identification experiments separated the effects of auditory qualities from categories by varying the acoustics of the target sound's context continuously rather than categorically, and having the listeners identify the context as well as the target. If listeners' categorization of the context influences their identification of the target independently from the context's acoustics, then processing is at first linguistically naïve and autonomous. The discrimination experiments compared listeners' ability to distinguish stimulus pairs in which the values of the manipulated acoustic property changed between context and target *within* a stimulus, H(igh)-L(ow) vs L-H, with those in which it only changed *between* stimuli, H-H vs L-L pairs. If listeners discriminate these pairs in terms of their auditory qualities, then contrast between the two intervals in the H-L and L-H pairs should make them more discriminable, but if they instead discriminate them in terms of categories, then the two kinds of pairs should be equally discriminable because the two intervals differ equally in their constituent categories.

One stimulus set consisted of a 7-step [l-r] continuum followed by a 7-step [d-g] continuum (LRDG stimuli) and the other of a 7-step [s-ʃ] continuum followed by a 7-step [t-k] continuum (SSHTK stimuli). Parameter manipulation created sounds that varied in whether energy was concentrated high in the spectrum, the [l,s,d,t] continuum endpoints, or low, the [r,ʃ,g,k] endpoints. The 49 consonant clusters in both sets were presented in [a__a] contexts. Prior work has shown that the listeners respond "g" more often after [l] than [r] and "k" more often after [s] than [ʃ] (Mann, 1980; Mann & Repp, 1981; Repp & Mann, 1981, 1982), i.e. with the category corresponding to the spectrally low endpoint in a spectrally high context. In other words, the target *contrasts* perceptually with its context.

Fig. 1: Effect of Liquid Acoustics on Stop Response

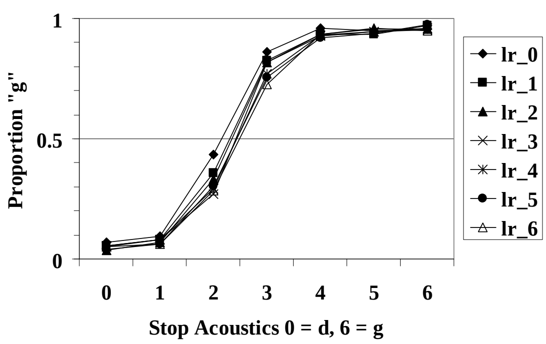
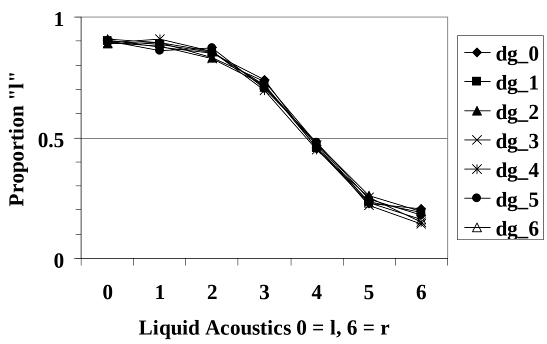


Fig. 2: Effect of Stop Acoustics on Liquid Response



Listeners responded "g" significantly more often as the preceding liquid became more [l]-like (Fig. 1; $F(6,114) = 7.405$, $p < .001$), but "l" responses were not affected by the following stop's acoustics (Fig. 2; $F(6,114) = 1.901$, $p = .087$). Stop and liquid identification did not significantly affect the other segment's identification (liquid response on stop response: $F < 1$; stop response on liquid response: $F(1,19) = 1.914$, $p > .10$), but their identification did interact significantly with the effect of their acoustics: for stimuli with ambiguous values along the [l-r] continuum, listeners responded "g" significantly more often when they identified the liquid as "l" ($t(22) = 3.863$, $p = .001$), and similarly for stimuli with ambiguous values along the [d-g] continuum, they responded "l" more often when they identified the stop as "g" ($t(22) = 4.874$, $p < .001$). That is, the category to which the context was assigned influenced target identification when the context's auditory quality would not bias listeners one way or another. Liquid identification influenced stop identification in the same direction as its acoustics do, more "g" responses after liquids identified as "l" as well as those whose acoustics are more [l]-like, while stop identification influenced liquid identification even though its acoustics did not.

Fig. 3: Effect of Fricative Acoustics on Stop Response

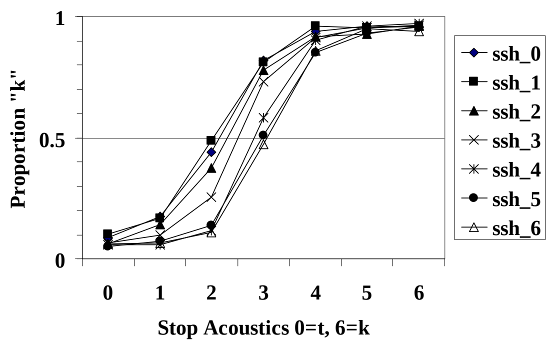
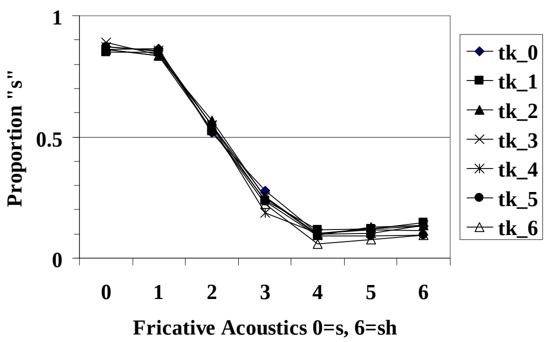


Fig. 4: Effect of Stop Acoustics on Fricative Response



Listeners responded "k" significantly more often as the preceding fricative became more [s]-like (Fig. 1; $F(6,84) = 56.248$, $p < .001$), and they responded "s" more often as the stop became more [t]-like (Fig, 4; $F(6,84) = 3.431$, $p = .003$). The fricative's effect on the stop's identification is contrastive, but the stop's effect on the fricative's identification is assimilative. The fricative response does not significantly influence "k" responses by itself ($F < 1$), but it interacts significantly with the fricative's acoustics ($F(6,84) = 3.284$, $p = .006$) because listeners responded "k" more often when they identified the fricative as "s" in the [s]-half of the continuum and when they identified the fricative as "ʃ" in the [ʃ]-half. Listeners also responded "s" significantly more often when they identified the stop as "k" rather than "t" ($F(1,13) = 51.641$, $p < .001$). The effect of the stop response is opposite that from its acoustics: listeners responded "s" more often when they identify the stop as "k" but also when the stop is acoustically more [t]-like.

The discrimination experiments presented 5 pairs of stimuli drawn from the quadrants and middle of the stimulus arrays. The pairs were either H-L vs L-H, i.e. [lg]- vs [rd]-like pairs or [sk]- vs [□t]-like pairs, or H-H vs L-L, i.e. [ld]- vs [rg]-like pairs or [st]- vs [□t]-like pairs. For the LRDG stimuli, listeners discriminated the H-L vs L-H pairs significantly better than the high-high vs low-low pairs (H-L vs L-H $d' = 2.467$, H-H vs L-L $d' = 2.173$; $F(1,30) = 7.751$, $p = .009$), but they did not also do so for the SSHTK stimuli (H-L vs L-H $d' = 3.016$, H-H vs L-L $d' = 3.115$; $F(1,29) = 1.571$, $p > .10$).

The autonomy of auditory processing from categorization is demonstrated by (1) the following stop's acoustics do not affect the identification of the preceding liquid but its identification does, (2) the following stop's acoustics affect identification of the preceding fricative in the opposite direction from its identification, and (3) liquid-stop sequences in which the two intervals may contrast auditorily are more discriminable than those in which they do not.

References

Mann, V. A. (1980). Influence of preceding liquid on stop-consonant perception. *Perception & Psychophysics*, 28, 407-412.

Mann, V. A., & Repp, B. H. (1981). Influence of preceding fricative on stop consonant perception. *Journal of the Acoustical Society of America*, 69, 548-558.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18, 1-86.

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*, 23, 299-370.

Repp, B. H., & Mann, V. A. (1981). Perceptual assessment of fricative-stop coarticulation. *Journal of the Acoustical Society of America*, 69, 1154-1163.

Repp, B. H., & Mann, V. A. (1982). Fricative-stop coarticulation: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, 71, 1562-1567.

The acquisition of lexical rhythm and duration by Japanese second language learners

Naoko Kinoshita* & Chris Sheppard#

*Integrated Education Learning Center, Meikai University, Japan; naoko-k@meikai.ac.jp
#Faculty of Science and Engineering, Waseda University, Japan; chris@waseda.jp

Introduction

Abercrombie (1967) hypothesized that language can be separated into syllable and stress timed languages. Subsequently, Han (1962) and Ladefoged (1975) suggested that mora timing be added as a further category. Although, several researchers believe that isochrony may not be a basis for categorizing language (Roach, 1982; Dauer, 1983), recent research has demonstrated that the variance in consonant and vowel length correlates with the predicted rhythmic classes (e.g., Ramus et al., 1999; Low, Grabe & Nolan, 2000; Grabe & Low, 2002).

Given the lack of status of the study of rhythm categories in language, and until recently, the lack of a valid measure of rhythm, there have been few studies which have examined the acquisition of rhythm by non-native speakers of the language or the influence of the rhythm of their first language. There have been even fewer in the case of the acquisition of Japanese as a second language. However, one such study is Toda's (2003) who determined that in addition to positive and negative transfer for English native speakers learning Japanese there was also an obvious development through the interlanguage towards the native speaker norm.

The present study first sets out to investigate the measurement of rhythm, and compares two new rhythmic measures, the pairwise variability index (PVI) developed by Grabe and colleagues (e.g. Grabe & Low, 2002) and the ΔV, ΔC and %V measures developed by Ramus et al. (1999) with the long/short ratios used in researching Japanese mora timing (e.g. Kashima, 1992).

Then the second, and main, research question aims to investigate the nature of productive rhythm development of Korean native speakers learning Japanese.

Methodology

For this study 10 Korean participants were selected from a group of volunteers based on the number of years they had spent studying the language. Five advanced learners (KH) who had learned Japanese at a Korean university for a minimum of three years were selected, and five beginning (KL) learners who had spent less than a year studying were also chosen. Five Japanese native speakers (NS) of the Tokyo dialect were also included in the sample.

Seven 2- and 3-syllable nonsense words were made up and used to elicit spoken samples from the participants (Mamaa, Maama, Mamaama, Maamamaa, Mamaamaa, Maamaama, and Maamaamaa). The words varied in the number of mora they contained. They were presented in *hiragana*, one of the Japanese scripts. Because the stimuli consisted of only two characters participants had sufficient control of the script.

After the participants had visually confirmed the stimulus, they were required to read each of the words twice aloud as they were presented on a computer screen. Their production was recorded with a Sony DAT and a condensing microphone, then uploaded into the computer and stored as 48 kHz 16 bit .wav files.

The intervocalic and the vocalic intervals for each of the sound files were then measured and recorded. These were used to produce three measures of alternating duration, as described above. The scores from the seven words were averaged for each participant. In order to respond to each of the research questions, first, correlations were calculated between the various measures.. To determine the rhythmic development of the learners, the duration measures of the learner groups were compared, both with each other and with those of the native Japanese using an ANOVA. Tukey's (HSD) post hoc tests were used for post hoc comparison in the event of significant difference being found. Probability was set at the five percent level for significance.

Results

The highest correlation was demonstrated between the vocalic nPVI and the ratio measure ($r = .87$, $p < 0.05$). The intervocalic PVI measure showed a negative moderate correlation ($r = -.61$, $p < 0.05$). Only one of Ramus's measures, ΔC, demonstrated a significant moderate correlation ($r = -58$, $p < 0.05$). The other measures of ΔV and %V failed to demonstrate any significant correlation with the ratio measure.

Based on the results of the correlation studies the vocalic nPVI and intervocalic rPVI measures were selected for subsequent analysis as these measures were judged to produce results comparable to those of previous research. The ANOVA demonstrated significant differences between the groups for the vocalic nPVI measure ($F = 3.25$, $p < .05$) only. Tukey's HSD determined that there was a significant difference between the KL and the other two groups (NS and KH).

Discussion

The correlations demonstrated that the PVI measures related best to the ratio measures used by researchers in Japanese in the past. This result is not surprising given the fact that both measures are comparing the relative length of adjacent vowels. Thus, the field now has a valid measure which can be extended beyond single vowel comparisons.

The second research question demonstrates that inexperienced Korean learners of Japanese develop towards the Japanese rhythmical norm as their interlanguage develops. This is supported by the fact that while the vocalic nPVI of the beginning learners was significantly different from NS, that of the advanced learners was not.

Conclusion

Overall it appears that the correlations demonstrated the PVI measure to be the better measure to use to compare with the ratio measures used by researchers in the past. Whether these measures are relevant to the measurement of both the syllabic nasal and the geminate consonant is yet to be seen.

The second main result demonstrated that the Japanese rhythm of Korean Learners develops towards the Japanese norm over time. This replicates the results of Toda (2003).

References

Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh: Edinburgh University Press.

Beckman, M. (1982). Segment Duration and the 'Mora' in Japanese. *Phonetica*, 39, 113-135.

Dauer, R. (1983). Stress-timing and syllable-timing re-analyzed. *Journal of Phonetics*, 11, 51-62.

Kashima, T. (1992). Nihongo no rizumu tanni to sono kata ni tsuite: Nihongo kyouiku heno ouyou wo mezashite. *Nihongo kennkyuu to nihonngo kyouiku*, Nagoya University, 305-319.

Grabe. E., & Low, E. (2002). Durational Variability in Speech and the Rhythm Class Hypothesis. In C. Gussenhoven & N. Warner (Eds.), *Papers in Laboratory Phonology 7* (pp. 377–401). Berlin: Mouton de Gruyter.

Han, M. (1962). The feature of duration of Japanese. *Study of Sounds*, 10, 65-75.

Homma, Y. (1981). Durational Relationship Between Japanese Stops and Vowels. *Journal of Phonetics*, 9, 273-281.

Ladefoged, P. (1975). *A Course in Phonetics*. New York: Harcourt Brace Jovanovich.

Low, E., Grabe, E., & Nolan, F. (2000). Quantitative characterizations of speech rhythm: 'syllable-timing' in Singapore English. *Language and Speech*, 43, 377-401.

Pike, K. (1945). *The Intonation of American English*. 2nd edition. Ann Arbor: University of Michigan Press.

Roach, P. (1982). On the distinction between 'stress-timed' and 'syllable-timed' languages. In D. Crystal (Ed.) *Linguistic Controversies* (pp. 73-79). London: Edward Arnold.

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73, 265-292.

Toda, T. (2003). *Second Language Speech Perception and Production: Acquisition of Phonological Contrasts in Japanese*. University Press of America.

An investigation of unaspirated dental and retroflex stops in Bangla

Aditi Lahiri* & Henning Reetz#

*Department of Linguistics and Philology, University of Oxford, UK, aditi.lahiri@ling-phil.ox.ac.uk
#Insitute for Phonetics, University of Frankfurt, Germany, reetz@em.uni-frankfurt.de

The lexical representation and processing of contrastive dental and retroflex place of articulation in Bangla is the core issue of the mediated priming study presented in this paper. The major outcome of this study is that retroflex sounds are the coronal defaults whereas dentals do not have an acoustic coronal property.

There is no real consensus in feature descriptions distinguishing a retroflex from a dental. In SPE (Chomsky & Halle, 1968), a retroflex was labelled as [-distributed, -anterior] while a dental was [+distributed, +anterior] to indicate that the retroflex was pronounced further back and involved the tip of the tongue. Bhat's (1974) description of retroflexion suggest that dentals become retroflexes often in the context of the high back vowel [u]. Panini's *ruki*-rule of retroflexion expresses that a dental [s] becomes a retroflex [ʂ] when adjacent to an [r u k] or [i], a group of sounds that hardly seem to have one feature in common. Surely it cannot be a front/back distinction because both [u] and [i] trigger the retroflexion. One possibility that fits into both Bhat's generalisation as well as the *ruki*-rule is that the context could well be described as [HIGH], while dentals are [LOW] (and both being [CORONAL]).

Stevens & Blumstein (1975) investigated the acoustic characteristics of dental, retroflex and velar unaspirated voiceless stops ([t̪, ʈ, k]) in Hindi. All have a falling F2, but dental stops have also falling F3 and F4, while retroflex and velar stops both have rising F3 and F4. Additionally, the burst energy is higher for the dentals (around 4 kHz) than for the retroflex (above F3) and lower for velars (between F2 and F3).

In a cross-modal priming experiment with 60 subjects and 240 target words with voiced and voiceless labial, dental, retroflex and velar stops we found in a direct semantic priming paradigm (e.g. the spoken word [din] played prior to the orthographic presentation of the word "sɔptaho") that every condition primed each other, i.e. there was no specific non-activation or inhibition in comparison to a control condition. Obviously, the semantic predominance overrides any subtle phonologic (de-)activation. In the more indirect 'mediated' paradigm (e.g. the spoken non-word [*gin] could activate the word [din], which itself is not presented to the participants, but in turn activates the semantically related word "sɔptaho", which is presented as visual target) we found the following pattern of activation and non-activation (milliseconds in brackets are difference to the control condition):

dental	->	velar	activation	(45 ms,	p<0.002)
velar	->	dental	non-activation	(-28 ms,	p<0.027)
retroflex	->	velar	non-activation	(0 ms,	p<0.999)
velar	->	retroflex	activation	(36 ms,	p<0.014)
velar	->	labial	non-activation	(18 ms,	p<0.204)
labial	->	velar	non-activation	(21ms,	p<0.175)

The results for velar and labial stops are no surprise, since the diffuse labial and compact velar spectrum are very different. The case for the dental and retroflex conditions in relation to the velar seems at first glance more puzzling (we left out the complete combinatorial set with the labials to make the experiment not to long for the participants). Especially the anti-symmetric behaviour of dentals and retroflex with respect to velars is striking (dentals activate velars, but not vice versa, and exactly the opposite for retroflexes). This is a strong indication that dentals and retroflexes are represented rather differently and are not two closely related 'sisters'. In acoustic terms, it is hard to understand that the dentals with more high-frequency energy activate the velars (with more energy in the less high frequency range) but not the other way round, and that retroflexes and velars, who share the same F3, F4 directions, do not activate each other in both directions.

The fact that dentals seem to become retroflexes indicate that the latter are less specific (are the more ‘common’ sounds) despite the fact that they occur less often in the language (but note that foreign loans with alveolars are indiscriminately adapted with retroflexes, e.g. English [ʈ]ime). We take this together as an indication that the representation of the segments is not primarily guided by acoustic properties but rather grounded on a more phonological representation, whose exact nature has to be further investigated.

Acknowledgements

The research was supported by the Leibniz Prize of the Deutsche Forschungs-gemeinschaft (DFG) to Aditi Lahiri.

References

Bhat, D. N. S. (1974). Retroflexion and retraction. *Journal of Phonetics*, 2, 233-237.

Chomsky, N., & Halle, M. (1968). *The Sound Pattern of English*, New York: Harper & Row.

Stevens, K. N., & Blumstein, S. E. (1975). Quantal aspects of consonant production and perception: a study of retroflex stop consonant. *Journal of Phonetics*, 3, 215-233.

Adult Acoustics and Developmental Patterns for Gender-marked Phonetic Variants in Mandarin Fricatives and Japanese Stops

Fangfang Li*, Eunjong Kong *, Jan Edwards #, & Mary E. Beckman *

* cathy; ejkong; mbeckman @ ling.ohio-state.edu
\# jedwards2@wisc.edu

Two questions of interest are addressed in the current study. First, how is social indexical information about gender represented in the adult acoustic space of a phonological category? Second, how does the acquisition of social indexical knowledge interact with the acquisition of lexical phonetic knowledge? In order to address these two questions, this study examines the acoustics of female-marked phonetic variants of obstruents of Songyuan Mandarin Chinese and Tokyo Japanese. In addition, to determine the relationship between the acquisition of this socially conditioned phonetic variation and the mastery of the related phonological contrast, we also compared productions by children with those by adult speakers in a database of target consonant productions elicited word-initially in varying vowel contexts.

The relevant Mandarin Chinese obstruents are the three voiceless sibilant fricatives, which contrast lexically in place of articulation and tongue posture: dental /s/, alveolopalatal /ɕ/ and retroflex /ʂ/. Hu (1991) and others have established that some Beijing dialect speakers front the alveolopalatal fricative to [sʲ], and that use of this variant is highest among young women. In the current study of the Songyuan dialect, native speaker transcriptions identified the same "feminine accent" variant in 3 out of the 10 women recorded. Acoustic analysis (Fig. 1) showed their alveolopalatal fricative productions to have centroid frequencies close to those for dentals, suggesting a constriction position as front as dentals for these talkers. The distinction between fronted alveolopalatals and dentals, however, was preserved in the second formant frequency at the onset of the following vowel, where higher onset F2 values for fronted alveolopalatals suggest a palatal constriction gesture, resulting in a shorter back cavity. Moreover, even the 7 women whose fricative productions were not transcribed as [sʲ] had higher relative centroid values for /ɕ/ than did the men. When the same acoustic analysis was used to describe children's productions, the results suggested that girls can begin to show fronting tendencies as early as age 4 or 5. Indeed, four out of the ten 4-year old girls and one of the 5-year-old girls produced extremely dentalized fricatives, which were transcribed as the feminine accent variant. It is worth noting that all children who produced the extremely dentalized fricatives were at least 4 years of age, and they produced the other two fricative categories with 100% accuracy, suggesting that acquisition of the social-indexical marker is dependent on a child's mastery of the lexical categories.

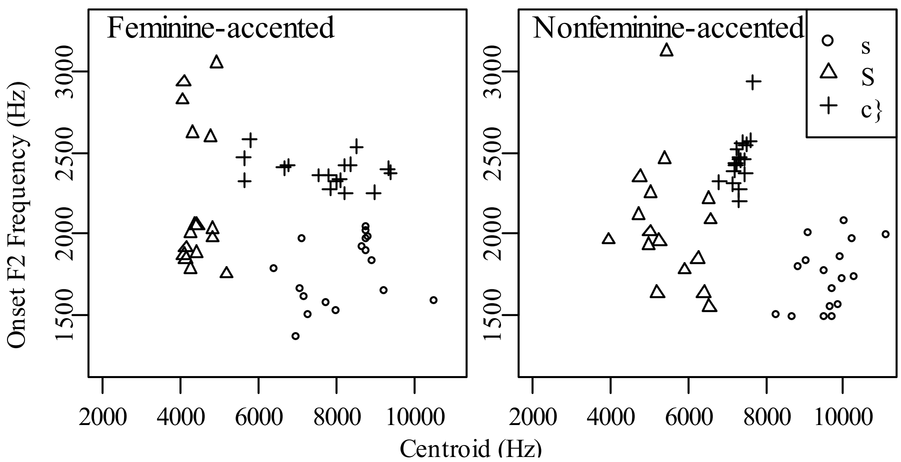


Figure 1: Acoustics of fricatives produced by representative women with and without "feminine accent".

The relevant Japanese contrast is the one between voiced and voiceless stops. Although earlier acoustic studies depict Japanese voiced stops as having voicing lead (e.g., Homma 1980), it has been recently noted that voiced stops vary along the VOT continuum and this variation is correlated with the speaker's age and gender (Takada

2004). Specifically, younger adult female speakers often produce voiced stops with a short lag VOT. The current study confirms this gender-related pattern for the 20 adults; 7 out of 10 women used short lag VOT voiced stops whereas only 2 out of 10 men did so (Fig. 2). The voiceless stops in these 9 speakers also showed a breathy voice quality at the following vowel onset, as indicated by the amplitude difference between the first and the second harmonics (H1-H2). By contrast, the Japanese-acquiring children did not show any clear gender-related differences in the VOT values for their voiced stops. Instead, nearly all of the 40 children examined (aged from 2 to 5 with ten for each age group) produce their voiced stops with short lag VOT values, regardless of gender. One interpretation of this result is that women (and children) are leaders in a sound change that will affect all speakers. However, it is also possible that children make the voiceless variant because truly voiced obstruents are difficult, as suggested by the very late acquisition of stops with voicing lead in other languages such as French, Hindi and Thai.

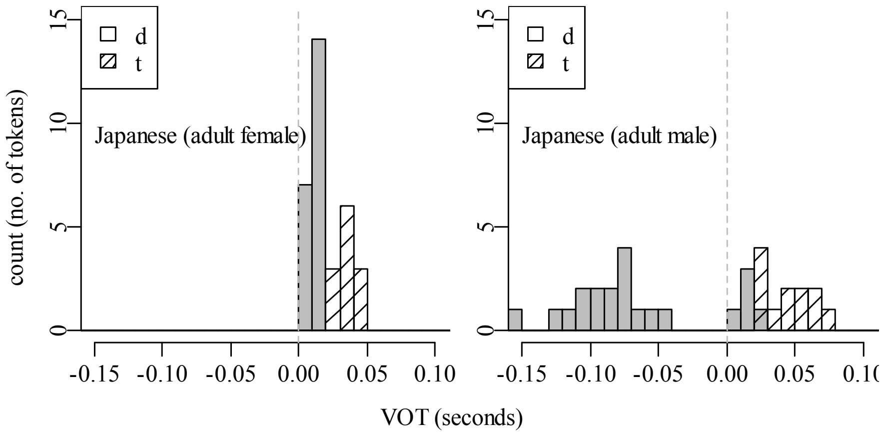


Figure 2: VOT values of Japanese stops produced by representative adult female (left) and male speaker (right).

The comparison of these results from the two languages prompts two thoughts about the relationship between social indexical gender specifications and the course of consonant phoneme mastery. First, the acoustic manifestations of the female variant seem to mimic the effects of diminutive size or childlikeness. Fronting an alveolopalatal fricative to make a higher centroid value mimics the effect of having a smaller vocal tract. Similarly, the voiceless variants of voiced stops are characteristic of canonical babbling and very early words that are produced concurrently with late babbling. Second, it appears that emergence of gender-marking variation could be constrained by the same maturational factors that govern the mastery of the lexical phonological contrast. Those Mandarin 4 or 5 year olds who exhibit extreme dentalized form in realizing feminine accent in their fricative productions all showed good proficiency in fricative production for all three lexically contrasting categories. By contrast, the Japanese children as a group might not have been able to show evidence of acquiring the voiceless variants as sociolinguistic markers of femaleness because the male-marking variant is a difficult sound that is not mastered until age 6 cross-linguistically. More generally, the acoustic evidence from the current study suggests that children's productions of gender-marking variants must be interpreted against that backdrop of both the language-specific phonological inventory and cross-linguistic developmental universals.

Acknowledgements

Work supported by NIH grant 02932 to Jan Edwards and an OSU research grant to the first two authors.

References

Homma, Y. (1980). Voice onset time in Japanese stops. *Onseigakkai Kaihoo*, 163, 7-9.

Hu, Mingyang (1991). Feminine accent in the Beijing vernacular: a sociolinguistic investigation. *Journal of the Chinese Language Teachers Association*, XXVI.1: 49-54.

Takada, M. (2004). VOT tendency in the initial voiced alveolar plosive /d/ in Japanese and the speakers' age. *Journal of the Phonetic Society of Japan*, 8(3), 57-66.

Phonetic Evidence for Neutralisation of Prelateral /æ/ and /e/ in Melbourne English

Deborah Loakes

School of Languages and Linguistics, The University of Melbourne, Australia; dloakes@unimelb.edu.au

Neutralisation involves the loss of a phonological contrast under certain conditions (see for example Trubetzkoy, 1939), a commonly cited example being loss of the voicing contrast syllable finally in German obstruents. There are a number of different ways in which neutralization may occur, and the focus of this paper is "Type A neutralization", where surface forms cannot be distinguished acoustically or perceptually (Dinnsen, 1985).

In this paper I investigate the apparent loss of a vocalic contrast between prelateral /æ/ and /e/ in Melbourne English, such that /e/ is realised phonetically as [æ] in this environment. In other words, minimal pairs in other varieties of English such as *Ellen-Alan* and *celery-salary* are not distinguished by speakers. This is precisely the phenomenon reported in New Zealand English (see for example Maclagan, 1982; Buchanan, 2001; Thomas & Hay, 2005).

Very little phonetic research has been carried out into this phenomenon in Australian English, but popular perception, and recent acoustic-phonetic research (e.g. Cox & Palethorpe, 2004), suggests that this neutralisation is confined to speakers from Victoria (also see Bradley, 1989). In this paper, I present acoustic-phonetic evidence for phonological neutralisation of prelateral /æ/ and /e/ in Melbourne English. This is based on a corpus of spontaneous speech produced by eight young male adults in two separate recording conditions. Vowels in words containing /æl/ and /el/ are analysed at the F1/F2 midpoint and compared with baseline /æ/ and /e/ values measured from controlled /hVd/ frames. Some speakers have disparate vowels in /el/ and /æl/ contexts, while others neutralise them (that is, /æ/ and /e/ are acoustically distinct in the /hVd/ frame, but the same before /l/). It should be noted that while this study reports results from spontaneous speech, Cox and Palethorpe (2004) note the same kind of prelateral neutralization in (Victorian female speakers') read speech. Additionally, results for those who neutralise prelateral /æ/ and /e/ agree with those observed for New Zealand English (e.g. Maclagan, 1982; Buchanan, 2001; Thomas & Hay, 2005).

As well as this phonetic evidence, I discuss neutralisation with respect to phonological theory, specifically exemplar theory (for example Pierrehumbert, 2001). With respect to this study, neutralisation is interesting for phonological theory because: a) Complete neutralisation occurs for some speakers and not others; b) It is not reported to occur anywhere in Australia besides Victoria (that is, it appears to be regionally specific); and c) Like Kim and Jongman's (1996) study of Korean manner neutralisation, the prelateral vowel neutralisation occurs despite potential cues for an underlying contrast in the orthography.

This study shows that speakers from the same region appear to use different phonological rules where prelateral /æ/ and /e/ are concerned (i.e. some neutralise the vowels, and some do not). Therefore, it is possible that speakers of Melbourne English have multiple underlying forms in the lexicon where prelateral /æ/ and /e/ are concerned. Conversely, exemplar clouds might simply have greater overlap for prelateral /æ/ and /e/ compared with the same vowels in non-prelateral environments. This latter idea agrees with discussion of the phenomenon in New Zealand English (e.g. Thomas, 2004), for which a greater amount of research has been carried out (including perception tests).

Finally, in this study the word *well* appears to pattern differently in comparison with other /el/ tokens, both for speakers who neutralise prelateral /æ/ and /e/ and those who do not, and so possible frequency effects for this item will also be discussed.

References

Bradley, D. (1989). Regional Dialects in Australian English Phonology. In P. Collins & D. Blair (Eds.), *Australian English: The Language of a New Society* (pp. 260-270). Brisbane: University of Queensland Press.

Buchanan, H. (2001). Neutralisation of DRESS and TRAP before /l/ in New Zealand English. *Wellington Working Papers in Linguistics*, 13, 15-29.

Cox, F., & Palethorpe, S. (2004). The Border Effect: Vowel Differences Across the NSW / Victorian Border. In C. Moskovsky (Ed.) *Proceedings of the 2003 Conference of the Australian Linguistics Society* (pp. 1-14).

Dinnsen, D.A. (1985). A Re-Examination of Phonological Neutralization. *Journal of Linguistics*, 21, 265-279.

Kim, H., & Jongman, A. (1996). Acoustic and Perceptual Evidence for Complete Neutralization of manner of Articulation in Korean. *Journal of Phonetics*, 24, 295-312.

Maclagan, M. (1982). An Acoustic Study of New Zealand English Vowels. *The New Zealand Speech Therapists' Journal*, 37, 20-26.

Pierrehumbert, J. (2001). Exemplar dynamics: Word frequency, lenition, and contrast. In J. Bybee & P. Hopper (Eds.), *Frequency effects and the emergence of lexical structure* (pp. 137-157). John Benjamins: Amsterdam.

Thomas, B. (2004). *In Support of an Exemplar-Based Approach to Speech Perception and Production: A Case Study on the Merging of Pre-Lateral DRESS and TRAP in New Zealand English.* Unpublished Masters Thesis, University of Canterbury.

Thomas, B., & Hay, J. (2005). A Pleasant Malady: The Ellen/Allan merger in New Zealand English. *Te Reo*, 48, 69-94.

Trubetzkoy, N.S. (1939). *Grundzüge der Phonologie*. Travaux du Cercle Linguistique de Prague 7.

A sociophonetic analysis of perception of sexual orientation in Puerto Rican Spanish

Sara L. Mack

Department of Spanish and Portuguese Studies, University of Minnesota – Twin Cities, USA; mack0185@umn.edu

A growing number of studies have explored possible links between phonological variation and listener evaluations of speaker sexual orientation in men's speech. These investigations have examined the influence of acoustic factors such as pitch range and pitch variability (Gaudio, 1994; Smyth et al., 2003), production of sibilants (Munson et al., 2006; Smyth et al., 2003), and various aspects of vowel production including both average F1 and F2 frequencies, and overall vowel-space expansion (Munson et al., 2006; Pierrehumbert et al., 2004; Podesva et al., 2001) on listener perceptions of speaker sexual orientation. In addition, studies have investigated relationships between perceptions of sexual orientation and perceptions of other speaker characteristics such as height, speech clarity (Babel & Johnson, 2006; Munson et al., 2006), perceived personality characteristics (Gaudio, 1994; Levon, 2006) and speech style (Smyth et al., 2003). However, there has been little systematic examination or documentation of such links in languages other than English. The current study contributes to this area of inquiry by exploring associations between phonetic variation and perception of sexual orientation in Spanish. This study, which is part of a larger project on the association between sociophonetic variation and voice recognition in Spanish and English, addresses three main questions: 1) Do Spanish-speaking listeners uniformly rate speakers' voices in terms of perceived sexual orientation? 2) If so, what phonetic variation correlates with different evaluations of perceived sexual orientation? And 3) What is the relationship between listener judgments of speaker sexual orientation and other perceived speaker characteristics?

The study was carried out in the San Juan, Puerto Rico, metropolitan area. Twenty different men were recorded saying a list of article-noun combinations. The men were recruited to include a variety of perceived sexual orientations. Eleven native speakers of Puerto Rican Spanish completed a computerized activity in which they listened to these different men's voices and reported their evaluations of four different speaker variables: age, height, social class, and sexual orientation. In each trial of the experiment, listeners heard six short phrases (*la gente, la manera, la vida, el tiempo, el día, el mundo*) produced by the same speaker. After listening to the phrases, listeners were presented with a question addressing one of the four variables. For age, listeners provided direct magnitude estimates of age in years. For height, social class, and perceived sexual orientation, listeners responded on a five-point equally appearing interval scale. The presentation order of speakers and questions was randomized.

The f0, F1, and F2 of the stressed vowels in each word, as well as the pretonic /a/ in *la manera* were measured using the Praat signal-processing software. Formant frequencies were measured in bark-scaled values, and f0 was measured in ERB. The F3 of the /a/ was also measured in Hertz. Average vowel-space expansion was calculated using the bark-scaled F1 and F2 values, based on the mean Euclidian distance formula first presented in Bradlow et al. (1996). Estimates of talkers' vocal-tract length were made by applying the F3 of /a/ to the odd-quarter length formula.

Analyses show that listeners do evaluate unique speakers' voices as different in terms of perceived sexual orientation. Perceptions of sexual orientation were most strongly predicted by average F2 frequency. Evaluations of the other speaker variables were most strongly predicted by different factors, or combinations of factors: perceived height by f0 and F2 frequency, and perceived social class by F1 frequency. An examination of the relationship between perceptions of sexual orientation and perceptions of height, age, and social class revealed that perceptions of height were correlated with perceived sexual orientation.

Put in broader context, the results of the study suggest that perception of sexual orientation based on phonetic variation is triggered by similar types of cues in Spanish and English. Furthermore, the fact that listener perceptions of speaker sexual orientation co-vary with listener notions of speaker age and height, even in short, audio-only stimuli, highlights the complex nature of social identities, as well as listeners' acute sensitivity to social information in speech.

References

Babel, M., & Johnson, K. (2006). Global talker characteristics and judgments of gay-sounding speech. Presentation given at the *35th Annual Meeting of New Ways of Analyzing Variation*, Columbus, OH, November 9–12.

Bradlow, A., Torretta, G., & Pisoni, D. (1996). Intelligibiilty of normal speech I: Global and finegrained acoustic-phonetic talkers characteristics. *Speech Communication*, 20, 255-272.

Gaudio, R. (1994). Sounding gay: Pitch properties in the speech of gay and straight men. *American Speech*, 69 (1), 30-57.

Levon, E. (2006). Hearing "gay": Prosody, interpretation, and the affective judgments of men's speech. *American Speech*, 81 (1), 56-78.

Munson, B., McDonald, E., DeBoe, N., & White, A. (2006). The acoustic and perceptual bases of judgments of women and men's sexual orientation from read speech. *Journal of Phonetics*, 34, 202-240.

Pierrehumbert, J., Bent, T., Bradlow, A., Munson, B., & Bailey, M. (2004). The influence of sexual orientation on vowel production. *Journal of the Acoustic Society of America*, 116 (4), 1905-1908.

Podesva, R., Roberts, S., & Campbell-Kibler, K. (2002). Sharing resources and indexing meaning the production of gay styles. In K. Campbell-Kibler, R. Podesva, S. Roberts, and A. Wong (Eds.), *Language and sexuality: contesting meaning in theory and practice* (pp. 175-189). Palo Alto, CA: CSLI Publications.

Smyth, R., Jacobs, G., & Rogers, H. (2003). Male voices and perceived sexual orientation: An experimental and theoretical approach. *Language in Society*, 32, 329-350.

Segmentation, not just similarity: alliteration as a natural experiment

Ian Maddieson

Department of Linguistics, University of New Mexico; ianm@berkeley.edu

Various forms of verbal art and verbal play provide illuminating evidence as to how speakers analyze the internal structure of words. For example, the use of end-rhyme is common in slogans, song lyrics and poetry in English (and a good number of other languages), as in (1)

(1) Chantilly lace and a pretty face
an' a pony tail a-hangin' down.
A wiggle an' a walk,
a giggle an' a talk
make the world go roun'. (J. P. Richardson "The Big Bopper", 1958)

Such data demonstrate that speakers make a division corresponding to the cut that linguists propose between the onset and rhyme of a syllable. Speakers are also aware of the location of stress, since all parts of the rhyming words after the onset of the stressed syllable must be identical for a true rhyme: 'funny' and 'money' rhyme with each other later in this text, but would not rhyme with 'knee'.

A different cut is demonstrated by the patterns of alliteration in verse and forms of ritual speech in early Germanic. These patterns provide strong *prima facie* evidence that speakers can be sensitive to the segment as a unit organizing their knowledge of the sound structure of their language. In essence the alliterative convention is as follows: one or two of the onsets of stressed syllables must begin with the same segment in two successive phrases (that is, poetic half-lines which generally contain two stresses), but not more than three onsets may alliterate in the line. Crucially, in the case of onsets containing clusters (except for */sp, st, sk/), only the first consonant is relevant. Early Germanic languages had a number of bisegmental onset clusters that have been simplified in some or all of the surviving languages, including */hw, hl, hr, hn, kn, wr, wl/ , in addition to many that have survived as clusters, such as */pr, pl, tr, kr, br, bl, dr, gr, gl, tw, dw, sw, sm, sn/ and so on. The initial consonant of all these bisegmental clusters alliterates with a matching single C onset (see 2, 3 below). Thus the identity of a single segment guides the pattern; the onset as a whole is not required to be identical, and the cut does not fall at the onset-rhyme boundary.

Although the most extensive data comes from manuscripts in Latin script from the 9th century C.E. or later, the tradition is not a convention based on literacy (Minkova, 2003; Lapidge, 1979). Early examples from Runic inscriptions testify to its antiquity. These include the inscription on the (now destroyed) shorter Golden Horn of Gallehus (Common Norse, 4th or 5th century C.E.) in (2) and the inscription on the front panel of the Franks Casket (Old English, early 7th century C.E.) in (3). Transcribed from the original runes into more familiar letters these read (with possible translations):

(2)	ek hlewagastiz holtijaz / horna tawido	"I, the ?honored-guest, Holti's son, made this horn"
(3)	fisc flodu / ahof on fergenberig	"the flood threw the fish onto the ?fir-mountain,
	warþ gasric grorn / þær he on greut giswom	the ghost-king was sad as he beached himself on the gravel"

In (2) the cluster /hl/ in 'hlewagastiz' alliterates with simple /h/ in 'holtijaz' and 'horna'. In (3) the cluster /fl/ in 'flodu' alliterates with simple /f/ in 'fisc' and 'fergenberig', and the cluster /gr/ in 'grorn' and 'greut' alliterates with simple /g/ in 'gasric' (though not with the 'g' in 'giswom' which is in an unstressed inflectional syllable).

The fact that this tradition provides support for the operational reality of a segment-sized unit seems to have attracted relatively little attention in the linguistic literature (or even for historical reconstruction of English prior to Minkova, 2003). This is in marked contrast to the substantial attention paid to two 'quirks' of the system — that clusters /sp, st, sk/ alliterate only with themselves but neither with simple /s/ nor other /s/-initial clusters, and that any vowel-initial word alliterates with any other vowel-initial word. The latter is probably because these words in fact regularly began with an unwritten glottal stop (Rapp, 1836; Minkova, 2003). The former remains potentially a challenge to the notion that segmentation is an explanatory account of the alliterative traditions.

One alternative was proposed by Fleischhacker (2000). In an experiment inspired in part by the patterns in the early Germanic alliterative tradition, she sought pairwise similarity ratings from speakers of modern American English. The subjects assigned similarity scores on a 7-point scale to rhyming word/non-word pairs such as [bleim]/[breim] (Cl/Cr), [bleim]/[beim] (Cl/C) or [breim]/[beim] (Cr/C). No significant difference was found between these three types of comparisons. That is, forms with an onset cluster containing a liquid were found to be perceptually equidistant both from each other and from a singleton onset. Fleischhacker therefore suggests that a generalized notion of degree of perceptual similarity can account for the acceptability of series such as /bl, br, b/ as good alliterating sets without appealing to segmentation (and presumably the same for sets such as /hr, hw, h/ and /kw, kn, k/), while /sp, st, sk/ are rejected as being too perceptually dissimilar.

A more elaborate version of this experiment is being run to see if the results are reproducible, to test the effects of task design on the results obtained, and to incorporate a wider range of onset types in the comparisons. Subjects are asked to perform two tasks. In one they provide similarity ratings of pairs of rhyming words using a 7-point scale, along the general lines of Fleischhacker's experiment. All pairs consist of either two real words of English or two nonsense words (e.g. 'fog/frog' or 'krig/kig'), rather than being word/non-word pairs. In the other task they judge which two out of a triplet of rhyming real (e.g. 'fog/frog/flog') or nonsense words (e.g. 'kig/krig/klig') are more similar to each other. The judgments from each task are obtained in two alternative ways labeled here 'written' and 'oral'. In the written presentation subjects read the words on a printed sheet and respond either by circling a number on the 7-point scale for the pairs task or by circling the two most similar words in the triplets task. In the oral presentation subjects are played a recorded voice speaking the pair or triplet of words to be judged (with a standardized falling pitch contour) and verbally report their judgment to the experimenter. Subjects perform a written task first, and an oral task second. Subjects who do the pairs task in written form do the triplets task orally, and vice-versa. Pilot results from 16 subjects have been analyzed to date, giving eight subjects in each task/presentation subset. Final reports aim to cover at least 40 subjects.

With respect to the pairwise comparison between singletons and obstruent+liquid clusters in the aggregated scores across subjects and presentations, Fleischhacker's result is approximately reproduced when raw rating scores are used. Cl/Cr pairs receive slightly higher ratings globally than C/Cr or C/Cl pairs, but the difference in ratings between these types of pairs does not reach a minimal level of significance ($p < .06$) in an analysis of variance. However, subjects vary considerably in the range of values on the scale which they use in reporting their comparisons, resulting in high inter-subject variance unrelated to the measurement of relative similarity. The effect of these differences can be removed by calculating the deviations from each individual subject's mean rating. When deviation scores are analyzed the main effect of pair type reaches the .01 level of significance. Cl/Cr pairs are rated as significantly more similar than pairs containing C/Cr or C/Cl. A difference also clearly emerges from the triplets task. In the triplets containing C/Cl/Cr onsets the Cl/Cr pairing is very strongly preferred as representing the two more closely similar words (66% of responses, against 33% chance). Neither the difference between real and nonsense words nor between written versus oral presentation has any consistent overall effect (though written vs oral presentation is confounded with subject differences, which can be quite marked). Taken together these results strongly suggest that C/Cl/Cr onsets are not all equally perceptually similar, as Fleischhacker had suggested. Rather, Cl and Cr onsets are more similar to each other than to singleton C onsets. If this intuition can be projected back to early Germanic times, then the alliterative patterns in early verse cannot be accounted for by onset-level similarity as this would suggest that a preference for Cl/Cr alliterative pairings over C/Cr or C/Cl should be found. Another factor must be in play. The obvious candidate is that similarity for the purposes of alliteration is only computed over the first segment of an onset for the majority of onsets.

The experiment also includes comparison of singleton /s/ vs s+stop onsets (sT) among a variety of other patterns. In the pairs test s/sT pairs are rated significantly less similar than either C/Cr ($p < 0.007$) or C/Cl pairs ($p < 0.003$). In the triplets task comparing s/sT/T onsets (e.g. 'seam, steam, team') the sT/T pair is judged most similar (60% of responses). This suggests that for modern English speakers the stop is the most salient element in the complex sT onset. If this intuition can also be projected back, it may help to account for why singleton /s/ onsets were not paired with sT clusters in the alliterative tradition. Alliterating /sp, st, sk/ only with themselves maintains the identity requirement on the initial segment and at the same time respects the primary salience of the stop element in these clusters.

References

Fleischhacker, H. (2000). *The location of epenthetic vowels with respect to consonant clusters: an auditory similarity account*. M.A. thesis. University of California, Los Angeles.

Lapidge, M. (1979). Aldhelm's Latin poetry and Old English verse. *Comparative Literature*, 31, 209-231.

Minkova, D. (2003). *Alliteration and sound change in early English*. Cambridge: Cambridge University Press

Rapp, K. M. (1836). *Versuch einer Physiologie der Sprache nebst historischer Entwickelung der abendländischen Idiome nach physiologischen Grundsätzen, erster Band*. Stuttgart: Cottasche Verlagsbuchhandlung.

Subglottal Resonances and Vowel Formant Variability: A Case Study of High German Monophthongs and Swabian Diphthongs

Andreas Madsack*, Steven M. Lulich#, Wolfgang Wokurek* & Grzegorz Dogil*

*University of Stuttgart - Institute for Natural Language Processing, Germany;
{andreas.madsack, wokurek, dogil}@ims.uni-stuttgart.de
#Massachusetts Institute of Technology – Speech Communication Group, USA; lulich@speech.mit.edu

Recent studies have shown that subglottal resonances can cause discontinuities in formant trajectories (Chi & Sonderegger, 2007), are salient in speech perception (Lulich, Bachrach & Malyska, 2007), and useful in speaker normalization (Wang, Alwan & Lulich, 2008), suggesting that variability in the spectral characteristics of speech is constrained in ways not previously noticed. Specifically, it is argued that 1) for the same sound produced in different contexts or at different times, formants are free to vary, but only within frequency bands that are defined by the subglottal resonances; and 2) for sounds which differ by certain distinctive features, certain formants must be in different frequency bands. For instance, given several productions of the front vowel [æ], the second formant (F2) is free to vary only within the band between the second and third subglottal resonances (Sg2 and Sg3), but in the back vowel [a] F2 must lie between the first and second subglottal resonances (Sg1 and Sg2). The feature [+/-back] is therefore thought to be defined by whether F2 is below Sg2 ([+back]) or above it ([-back]; Stevens, 1998). This has been tested mostly in English, and only recently in Korean (Jung, 2008). We present new evidence that the same definition applies to both High German and Swabian German monophthongs, and suggest that it can also account for the contrast between the Swabian German diphthongs [aj] and [əj].

We made standard microphone recordings of twelve native speakers of German, including eight Swabians, and four fluent non-native speakers (whose native languages were Russian (1), Georgian (2), and Turkish (1)). The speakers read sentences which included both of the [aj] and [əj] diphthongs (non-Swabians produced them without a distinction), as well as sentences containing nonsense words designed to elicit monophthongs in a neutral phonetic environment. In the latter case, the carrier sentence was ‘Peter hat hVd gesagt’ (‘Peter said hVd’), where the vowel was any of the standard German monophthongs. We also made recordings from an accelerometer placed on the skin of the neck below the larynx, which measured subglottal acoustics.

Front vowels, including the rounded front vowels, had F2>Sg2, and back vowels had F2<Sg2. There was some variability with regard to the low back vowel [a], for which the relative frequencies of F2 and Sg2 were dependent on the speaker. The low vowel [a] generally had F1>Sg1 whereas all other vowels had F1<Sg1. This is congruent with the hypothesis in Stevens (1998) that the feature [+/- low] is defined by a relation between Sg1 and F1 (see Table 1).

Table 1: Percentage of monophthong tokens obeying the subglottal hypothesis

Front vowels *(n=24),* F2>Sg2	96.88%	96.88%	97.92%
Back vowels *(n=18)*, F2<Sg2	62.50%	70.83%	61.11%
Low vowels *(n=6)*, F1>Sg1	72.92%	87.50%	83.33%
Non-low vowels *(n=36)*, F1<Sg1	90.63%	97.92%	98.61%

The Swabian diphthongs [aj] (e.g. ‘Weide’, ‘willow’) and [əj] (e.g. ‘Weide’, ‘pasture’) are an interesting case because no consistent spectral distinction between them has been found, and speakers of non-Swabian German dialects do not consistently perceive a difference between them. It has been suggested that different temporal patterns of F2 movement during the diphthongs underlie the contrast (Geumann, 1997; Hiller, 2003). We conducted a new analysis of the original speech data reported by Geumann (1997), as well as an analysis of tokens produced by our High German and Swabian German subjects. The High German speakers showed no systematic pattern to distinguish the two diphthongs, as expected. Of the six speakers in Geumann

Geumann (1997), there was a consistent spectral difference between the two diphthongs in all but Speaker 3: F2 in [aj] either began below Sg2 before crossing into higher frequencies, or it dipped below Sg2 briefly before crossing again into the higher frequencies (the latter case occurred when a consonant with a high F2 locus preceded the diphthong). In [əj], on the other hand, F2 began above Sg2 in almost all cases and rarely dipped below it (see Figure 1). Surprisingly, none of our Swabian German speakers showed a similar pattern. It appears, therefore, that specific sub-dialects within Swabian may distinguish these two diphthongs.

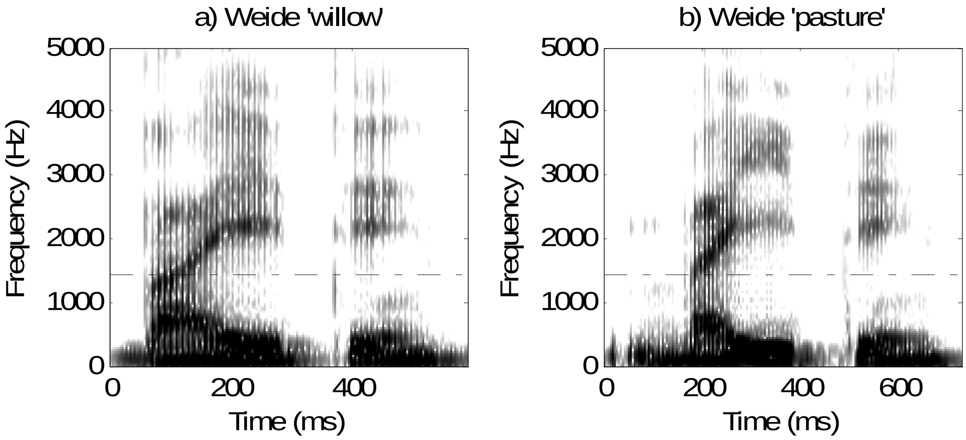


Figure 1: Spectrograms of a) 'Weide' ([ʋajdə], 'willow') and b) 'Weide' ([ʋəjdə], 'pasture'). The horizontal dashed lines mark the frequency of Sg2.

The results suggest that for Swabian German speakers who distinguish the [aj] and [əj] diphthongs, this contrast is accomplished by a spectral cue, namely, the frequency of F2 relative to Sg2 at the beginning of the diphthong. Previous findings that the difference between the diphthongs is temporal (Geumann, 1997; Hiller, 2003) must be revisited in light of our findings. It may be that the temporal differences are due to the interaction between F2 and Sg2, or that both spectral and temporal cues independently contribute to this contrast. In sum, these data support the hypothesis that speech spectral variability is constrained by subglottal articulatory-acoustic non-linearities which underlie distinctive features (Stevens, 1998).

Acknowledgments

This work is a part of SFB 732 and was partly funded by the German Research Association (DFG).

References

Chi, X., & Sonderegger, M. (2007). Subglottal coupling and its influence on vowel formants. *Journal of the Acoustical Society of America*, 122, 1735-1745.

Geumann, A. (1997) Formant trajectory dynamics in Swabian diphthongs. *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München*, 35, 35-38.

Hiller, M. (2003). The diphthong dynamics distinction in Swabian. In J. van de Weijer, V.J. van Heuven & H. van der Hulst (eds.) *The Phonological Spectrum*. Amsterdam: John Benjamins.

Jung, Y. (2008). Acoustic articulatory evidence for quantal vowel categories across languages. Poster presented at the Harvard-MIT HST Forum.

Lulich, S.M., Bachrach, A., & Malyska, N. (2007). A role for the second subglottal resonance in lexical access. *The Journal of the Acoustical Society of America*, 122, 2320-2327.

Stevens, K.N. (1998). *Acoustic Phonetics*. Cambridge, MA: MIT Press.

Wang, S., Alwan, A., & Lulich, S.M. (2008). Speaker normalization based on subglottal resonances. In *Proceedings of ICASSP-08* (pp. 4277-4280). Las Vegas, USA.

An investigation of cross-language differences in pitch range for speakers of English and German

Ineke Mennen*, Felix Schaeffler#, & Gerard Docherty^

*School of Linguistics, University of Bangor, UK; els803@bangor.ac.uk
#Speech Science Research Centre, Queen Margaret University, UK; fschaeffler@qmuc.ac.uk
^School of Education Communication & Language Sciences, Newcastle University, UK; g.j.docherty@ncl.ac.uk

A key issue for the Laboratory Phonology community is to understand along which phonetic dimensions languages can systematically differ, and how such differences are actually generated in speech production, evaluated in speech perception, and acquired by those learning to speak a language for the first time.

While there has been a good deal of effort devoted to understanding the intonational structure of speech and its relationship to the form and timing of salient aspects of an f0 contour across different languages (giving rise, of course, to many interesting discussions about the what those salient aspects might be), there has been much less focus on pitch range as a source of cross-language differences. For analysis purposes, pitch range can be divided into two sub-characteristics of a speaker's performance; pitch level and pitch span. While there are clearly organic and physiological factors which govern inter-individual variation in both of these quantities (e.g. greater vocal fold mass would, all else being equal, normally be associated with relatively low pitch levels), there is also some evidence that when groups of speakers of different languages are compared there can be a significant difference in aspects of pitch range notwithstanding a large degree of overlap in the range deployed by many speakers of both languages (i.e, a difference which is a characteristic of the collective but not *necessarily* of individual speakers).

Two languages where this may well be the case are Southern Standard British English (SSBE) and Northern Standard German (NSG). For example, there is strong anecdotal evidence that people perceive differences between speakers of English and German - with English sounding higher and having more pitch variation than German. British voices (especially female) are often perceived stereotypically as "over-excited" or even "aggressive" by German listeners. Conversely, to British listeners, German low-pitched voices may be evaluated as sounding "bored" or "unfriendly".

The aim of this study is to examine the extent to which this putative difference can be validated by a quantitative analysis of the pitch range of speakers of these two languages and, if it can be, to consider the implications for our understanding of language-specific phonetic specification.

There are two primary dimensions to our investigation. The first is to evaluate which measures are best suited to capture any cross-language differences in pitch range. The second is to use a range of these measures to attempt to identify the nature of the cross-language differences in performance which underpin the differences that people perceive. In pursuit of this second objective, we have carried out perceptual experimentation in order to verify which aspects of pitch range variation lead listeners to be more likely to identify a stimulus as being produced by a German speaker than by an English one.

Our experimental work is centred on the performance of a range of materials by 60 speakers, 30 for each of SSBE and NSG, all female university students in their twenties and thirties and functionally monolingual (i.e. no more than moderately proficient in another language). 30 participants took part in the perceptual experiments in which they were presented with resynthesised pitch contours from the production study and asked to judge how German or English the speaker sounds.

In this paper we present results from both aspects of this study. We first focus on a comparison of the findings from various measures derived from the raw f0 time series extracted from a stretch of speech and corrected for tracking errors (e.g. measures of mean/median f0, long term distributional measures such as max-min f0, 90% range) and measures derived from specific linguistically-relevant turning points in the f0 contour (as suggested in previous work by Patterson (2000) which in turn built on work by Ladd & Terken (1995) and Shriberg et al. (1996). In presenting the findings of our perception study, we highlight the correlation between listener judgements and our different pitch range measures.

In discussing our results, we highlight the advantages and disadvantages of the two types of measures for cross-language comparisons., we consider a number of factors which might account for listeners' sensitivity to global pitch range differences but which have not yet been embedded within our experiments (such as the time that speakers spend near the top or bottom of their range), and we provide a preliminary account of how such cross-language differences might come to exist in the first place and how they may be propagated across generations.

Acknowledgements

Project funded by UK Economic & Social Research Council [R000221858]. We thank Frank Kügler and his colleagues for the collection of the German data.

References

Ladd, D. R., Terken, J. 1995. Modelling Intra- and Inter-Speaker Pitch Range Variation. In *Proceedings of the 13th International Congress of Phonetic Sciences* (pp. 386-389). Stockholm, Sweden.

Patterson, D. (2000). *A linguistic approach to pitch range modelling.* PhD dissertation, University of Edinburgh.

Shriberg, E., Ladd, D.R., Terken, J.,& Stolcke, A. (1996). Modeling pitch range variation within and across speakers: predicting F0 targets when "speaking up". In *Proceedings for the International Conference on Spoken Language Processing* (Addendum, 1-4). Philadelphia, PA.

Syllabifications of the /st/ cluster and vowel-to-vowel coarticulation in English

Peggy Mok* & Sarah Hawkins#

*Department of Linguistics and Modern Languages, The Chinese University of Hong Kong; peggymok@cuhk.edu.hk
#Department of Linguistics, University of Cambridge, U.K.; sh110@cam.ac.uk

Introduction

This paper investigates how different syllable boundaries involving the /st/ cluster in English affect vowel-to-vowel (v-to-v) coarticulation. Although several models of coarticulation offer accounts of vowels and consonants in different syllable positions (e.g. Articulatory Phonology (Browman & Goldstein, 1988; 2000) and 'carrier' models of coarticulation (e.g. Öhman, 1966)), the effects of syllable structure on v-to-v coarticulation remain poorly understood. Most studies on v-to-v coarticulation only deal with one syllable type, namely open syllables (though see Modarresi, Sussman, Lindblom & Burlingame, 2004). However, notwithstanding the lack of a clear definition of a phonetic syllable, many studies show that syllable onset and coda are different acoustically, articulatorily, typologically and perceptually.

Acoustic studies indicate that onset consonants are longer and exhibit stronger cohesion with tautosyllabic vowels than do coda consonants (e.g. Byrd, 1996; Sussman, Bessell, Dalston & Majors, 1997). Articulatory studies show that syllable onset and coda consonants coordinate differently with the vowels, and that gestures for onset consonants are stronger and more distinct than those for coda consonants (e.g. Browman & Goldstein, 1988; Krakow, 1999). Onset consonants are more frequent in the world's languages, and are more distinguishable than coda consonants in noise (Redford & Diehl, 1999). All this suggests that syllable onset and coda consonants have different coordination with the vowel, which might therefore be expected to affect v-to-v coarticulation.

The /st/ cluster in English can be syllabified in three ways: onset /#st/, heterosyllabic /s#t/ and coda /st#/ (where # denotes a syllable boundary). It was hypothesized that the onset /#st/ should allow the least v-to-v coarticulation because onsets are stronger and more stable, followed by the heterosyllabic /s#t/ and the coda /st#/ because coda consonants are the most variable. The /st/ cluster was chosen for investigation because it is homorganic, thus reducing conflicting influences of intervocalic consonants on formant transitions.

Method

Six native speakers of Southern British English (two male four female) were recorded reading sequences of two real monosyllabic words involving the /st/ cluster embedded in carrier phrases. Three sets of materials were collected: onset /CV_1#stV_2C/ (e.g. bar steed), heterosyllabic /CV_1s#tV_2C/ (e.g. pass teat) and coda /CV_1st#V_2C/ (e.g. past east). Three vowels, /ɑ/, /i/ and /u/ were used in all possible combinations. Focus stress was induced on the non-target syllables by the carrier phrases (Not a <u>x x</u> , but a <u>x x</u> again.). Altogether 324 sequences were used (3 syllable structures × 3 target vowels × 3 context vowels × 2 stress positions × 6 repetitions). F1, F2 and F3 frequencies were measured from LPC spectra (25 ms hanning window, supplemented by DFT spectra) at several temporal locations: 1) close to the offset of periodicity in V_1 or onset of periodicity in V_2 for anticipatory and carryover coarticulation respectively; 2) 26 ms after the /t/ burst (during the aspiration), in order to measure carryover coarticulation at a comparable place in the articulatory trajectory during heterosyllabic /CV_1s#tV_2C/ sequences. All formant frequencies were normalised before being submitted to statistical analysis. Intervocalic durations were also measured.

Results and discussion

Results show that heterosyllabic /s#t/ sequences had the longest intervocalic duration presumably because they contain a syllable/word boundary. The intervocalic durations for onset /#st/ and coda /st#/ are generally not different from each other. However, four-way repeated measures ANOVAs (direction × syllable structure × target vowel × context vowel) of the normalised formant frequencies suggested that at the comparable place in the articulatory trajectory, different syllabifications of the /st/ cluster did not significantly affect the degree of v-to-v coarticulation, contrary to the hypothesis based on the various differences between onset and coda consonants.

Articulatory Phonology suggests that syllable structures result from different gestural overlap and timing patterns between vowels and consonants, with onset consonants showing the C-centre effect with the vowel while only the left-most consonant in the coda cluster is phased with respect to the vowel. This predicts that different syllabifications of the /st/ cluster should affect v-to-v coarticulation differently. However, the present results do not support this prediction. Rather, the results seem to be more compatible with 'carrier' models of coarticulation (e.g. Öhman 1966) in which vowels form a continuous diphthongal movement onto which consonantal gestures are superimposed, such that syllabification of the consonants does not affect coarticulation between successive vowels, although such models may not be applicable to all languages (see Smith, 1995).

References

Browman, C. P., & Goldstein, L. (1988). Some notes on syllable structure in Articulatory Phonology. *Phonetica,* 45, 140-155.

Browman, C. P., & Goldstein, L. (2000). Competing constraints on intergestural coordination and self-organisation of phonological structures. *Les Cahiers de l'ICP, Bulletin de la Communication Parlée,* 5, 25-34.

Byrd, D. (1996). Influences on articulatory timing in consonant sequences. *Journal of Phonetics,* 24, 209-244.

Krakow, R. A. (1999). Physiological organization of syllables: a review. *Journal of Phonetics,* 27, 23-54.

Modarresi, G., Sussman, H., Lindblom, B., & Burlingame, E. (2004). An acoustic analysis of the bidirectionality of coarticulation in VCV utterances. *Journal of Phonetics,* 32, 291-312.

Öhman, S. E. G. (1966). Coarticulation in VCV utterances: spectrographic measurements. *Journal of the Acoustical Society of America,* 39, 151-168.

Redford, M., & Diehl, R. (1999). The relative perceptibility of initial and final consonants in CVC syllables. *Journal of the Acoustical Society of America,* 106, 1555-1565.

Smith, C. L. (1995). Prosodic patterns in the coordination of vowel and consonant gestures. In B. Connell & A. Arvaniti (Eds.), *Papers in Laboratory Phonology IV: Phonology and Phonetic Evidence* (pp. 205-222). Cambridge: Cambridge University Press.

Sussman, H., Bessell, N., Dalston, E., & Majors, T. (1997). An investigation of stop place of articulation as a function of syllable position: a locus equation perspective. *Journal of the Acoustical Society of America,* 101, 2826-2838.

Integrating Sub-phonemic Cues: learning across a morpheme boundary

Rebecca Morley

Department of Cognitive Science, Johns Hopkins University, Baltimore, MD; morley@cogsci.jhu.edu

A basic idea, central to Evolutionary Phonology, is that the speech signal, produced with varying degrees of reduction and overlap of articulations, is vulnerable to being misparsed, and that this misparsing is the primary source of internal language change (Blevins, 2004). Ultimately, every perceivable acoustic difference has the potential to be realized, at some point in time, and in some language, as a phonological distinction. If this theory is correct, we should be able to show that common phonological distinctions possess consistent phonetic correlates. Furthermore, these phonetic precursors must have some way of interacting with the already existing grammar of a language in order to effect further possible changes. In terms of the type of system I will be most interested in here, namely, derived environment effects, there must be some acoustic way to encode a domain difference within and across morphemes.

This hypothesis was tested experimentally using an artificial grammar learning paradigm wherein participants were exposed to novel words in a made-up language. These words consisted of pairs of bare and inflected forms, where the morphologically complex member was constructed by the addition of the suffix /-m/ (e.g., [skimtu]; [skimtum]). This morphological alternation was associated with a phonetic cue (degree of regressive vowel nasalization, difficult to consciously discriminate for English speakers). High nasalization vowel tokens were created by splicing from a stressed nasal coda environment such that the entire vowel length was nasalized (second nasal formant visible throughout). Low nasalization tokens were created by splicing from an oral environment, such that nasalization was absent (no nasal formant visible in any part of the vowel).

In Condition A, participants were trained on words that exhibited high nasalization across the morpheme boundary, but low nasalization on pre-nasal vowels adjacent to a tautomorphemic nasal. In Condition B, these values were reversed, such that low nasalization occurred across the boundary, and high within. The testing procedure was a two alternative forced choice task, where the two auditorily presented words differed only in their degree of nasalization (and participants were asked which of the two was more likely to have been produced by the speaker they had previously listened to). Some of these words were familiar, and some were novel; words varied with respect to syllable structure, length and segment make-up. See Table 1 for example training and test items.

Table 1: Example training and test items

Condition	Train: Singular	Train: Plural	Test (Novel) : Singular	Test (Novel) : Plural
A	skimtu	skimtũm	ʒadimfu vs. ʒadĩmfu	ʒadimfũm vs. ʒadimfum
B	skĩmtu	skĩmtum	ʒadimfu vs. ʒadĩmfu	ʒadĩmfũm vs. ʒadĩmfum

A robust interaction effect (Condition [A/B] x Type [singular/plural]) for both old and new items indicates that learners were able to encode detailed phonetic representations (two differing degrees of nasal coarticulation within a given word), as well as make structural associations, enabling generalization to novel words. A consistent difference in accuracy between low and high nasalization items reveals an asymmetry such that learners were more permissive of low forms when they were trained on high; but were less likely to accept high forms when trained on low. This may be because high nasalization tokens show greater variability in actual degree of nasalization than their low counterparts, such that the “HI nasalization” condition is actually a “high variability” condition (work is currently underway to test this).

The experimental work presented here suggests that we can perceive and encode different phonetic associations for morphological boundary versus non-boundary environments. This provides support for a theory of rich lexical representations. This result is particularly compelling in the present case where participants might have been expected to disregard vowel quality differences that are never contrastive in their native language (or in the exposure artificial grammar). Furthermore, the fact that, typologically, both systems of over-application and under-application in derived environments are attested, is supported by the learnability of associations in both directions (condition A showed a slight numeric advantage in accuracy over condition B, but this effect did not reach significance). This might be characterized as a learned degree of coarticulation, or perhaps a degree of variability of coarticulation (see Cho, 2001). A naturalness bias might exist, based on universal properties of motor planning and articulation, but these preliminary results suggest that phonetic cues are more broadly available for implementation by a grammar, given enough statistical evidence. Finally, this result can be related to questions about the emergence of phonological constraints from a possible phonetic substrate. If the necessary condition of a lexical inventory of highly acoustically detailed representations is met, then generalization might proceed over that lexicon to produce a more abstract grammatical rule.

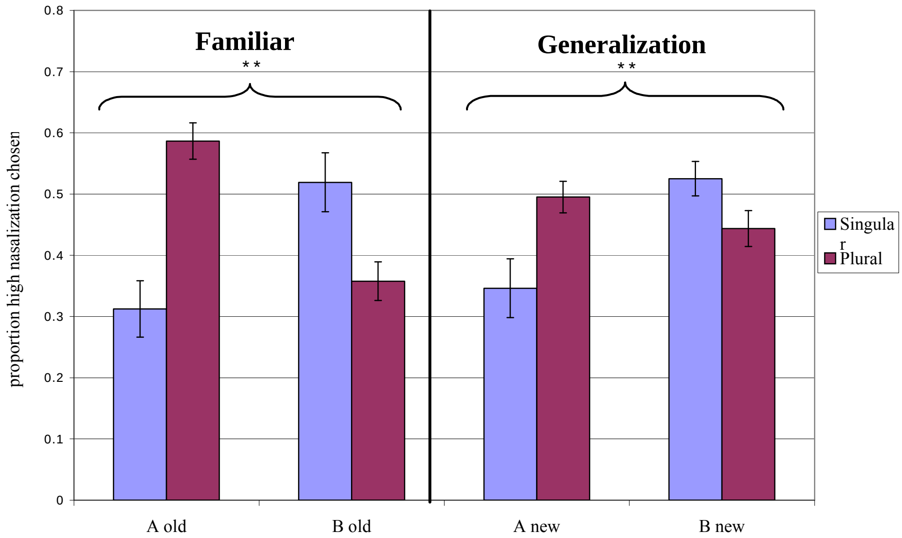


Figure 1: Proportion high nasalization chosen

Acknowledgments

This work was supported by an NSF IGERT training grant and a Department of Education Javits fellowship. I would especially like to thank Paul Smolensky, Ariel Goldberg and the members of the Johns Hopkins IGERT lab past and present.

References

Blevins, J. (2004). *Evolutionary Phonology: the emergence of sound patterns*. New York: Cambridge University Press.

Cho, T. (2001). Effects of morpheme boundaries on intergestural timing: evidence from Korean. *Phonetica*, 58, 129-162.

Articulatory Marking of Focus in German

Doris Muecke, Martine Grice, Anne Hermes, Johannes Becker & Stefan Baumann

I*f*L Phonetik, University of Cologne, Germany;
{doris.muecke; martine.grice; anne.hermes; becker.johannes; stefan.baumann}@uni-koeln.de

Introduction

This study reports on a production experiment investigating articulatory means of encoding different focus structures in German. Previous investigations in this field have been restricted to words in maximally diverging focus structures (contrastive focus vs. background) and thus to the accented-unaccented dichotomy (Cho, 2005; Avesani et al., 2007). It is thus unclear from these studies whether the articulatory differences found, such as greater jaw lowering or lip aperture in contrastive focus, are simply due to the presence of an accent, or whether they involve the expression of additional prominence or emphasis as well (Dohen et al., 2006), suggesting that differences in articulation might be found independently of whether a word is accented or not.

We concentrated on lip kinematics (Byrd, 2000) in CVC target syllables, comparing different types of focus (contrastive, non-contrastive) and different sizes of focus domain (broad, narrow), as well as comparing focus and background. The lip parameters investigated are greater displacement, longer duration, higher peak velocity and lower stiffness, all of which may enhance the sonority and the prominence of the vowel.

Methods and Results

Three native speakers of Standard German were recorded with a Carstens AG100 Articulograph, with sensors placed on the vermillion border on the upper and lower lip. Four disyllabic words were constructed in the frame *Doktor B/V:/ber*, where V: (/i:/, /a:/, /o:/ or /u:/) is the vowel in a fictitious surname. Each target name was produced in four focus structures (seven repetitions), making a total of 112 tokens per speaker.

Question-Answer-Pairs:	**test words in:**
1. Will Norbert Doktor Bahber treffen? *Does Norbert want to meet Dr. Bahber?*	
[Melanie] $_{focus}$ will Doktor Bahber treffen.	background
2. Was gibt´s Neues? *What´s new?*	
[Melanie will Doktor Bahber treffen]	broad focus
3. Wen will Melanie treffen? *Whom does Melanie want to meet?*	
Melanie will [Doktor Bahber] $_{focus}$ treffen.	narrow focus
4. Will Melanie Doktor Werner treffen? *Does Melanie want to meet Dr. Werner?*	
Melanie will [Doktor Bahber] $_{focus}$ treffen.	contrastive focus
(Answer lit.: *Melanie wants Dr. Bahber to meet*) Focus marked with [...] $_{focus}$	

Accent Types: Although all three speakers deaccented backgrounded items, only DM and AH used accent types to distinguish between different focus structures (Table1).

Table 1: Most frequently produced accent types per focus condition

speaker	background	broad focus	narrow focus	contrastive focus
DM	Ø	!H*	^H*	^H*
AH	Ø	!H*	!H* H* ^H*	^H*
WP	Ø	^H*	^H*	^H*

Articulation: Figure 1 provides averaged trajectories for the distance between the upper and lower lip (Lip Aperture) during the production of the target word B/i:/ber for speaker DM (left) and speaker AH (right). Going from background through broad and narrow to contrastive focus we found an increase in displacement, duration and peak velocities. Differences reached significance not only when comparing background to contrastive focus, a result which is to be expected, given the literature on other languages (Cho 2005; Avesani et al., 2007), but also when comparing broad and contrastive focus, where the target syllable is accented in both cases.

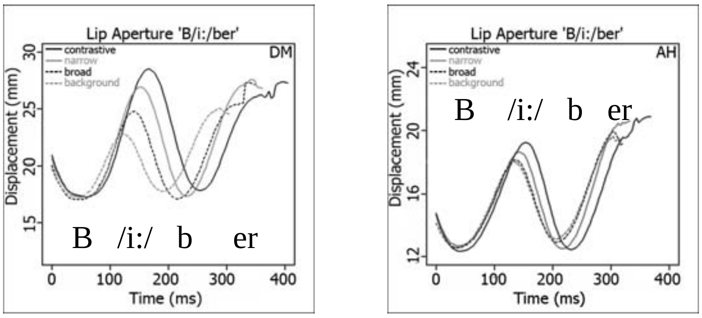


Figure 1: Averaged contours for Lip Aperture, speaker DM and AH, target word B/i:/ber

Figure 2 provides medians and quartiles for stiffness in the opening gesture for the speakers DM and AH, where stiffness is related to the relative speed of the articulatory movement (time-to-peak velocity provided similar results). Contrastive focus tokens have significantly lower stiffness than backgrounded ones, as found for English in Beckman et al. (1992). Our data also revealed significantly lower stiffness values in contrastive as opposed to broad focus contexts.

In a mass-spring model (Byrd, 2000), the articulatory adjustments can be accounted for by rescaling of the respective gestures which includes a change of the stiffness and target parameters. However, speakers behave differently: we found non-linear rescaling for DM and linear rescaling for AH.

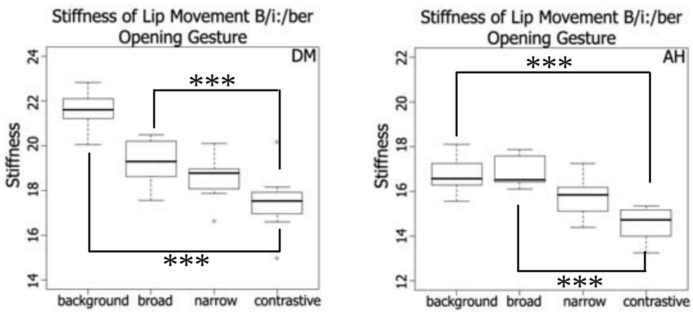


Figure 2: Stiffness (ratio of peak velocity to maximum displacement), speaker DM and AH

In sum, we found systematic differences in a number of articulatory parameters corresponding to sonority expansion between the marking of constituents as background on the one hand, and as contrastive focus on the other, in particular larger displacements and lower stiffness on the latter. We show that sonority expansion is not simply a concomitant of accentuation, since we also found comparable differences in articulatory adjustments when comparing broad and contrastive focus. Thus we show that speakers of German differentiate between linguistic categories relating to focus structure not only by means of accentuation, but also by varying aspects of their supralaryngeal articulation patterns resulting in the enhancement of the syntagmatic contrast between consonants and vowels on and around the target syllable.

References

Avesani, C., Vayra, M., & Zmarich, C. (2007). On the articulatory bases of prominence in Italian. In Jürgen Trouvain & William J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 981-4). Saarbrücken, Germany.

Beckman, M. E., Edwards, J., & Fletcher, J. (1992). Prosodic structure and tempo in a sonority model of articulatory dynamics. In G.J. Docherty & D.R. Ladd (Eds.), *Papers in Laboratory Phonology II* (pp. 68-86). Cambridge: Cambridge University Press.

Byrd, D. (2000). Articulatory vowel lengthening and coordination at phrasal junctures. *Phonetica*, 57, 3-16.

Cho, T. (2005). Prosodic strengthening and featural enhancement: Evidence from acoustic and articulatory realizations of /ɑ, i/ in English. *JASA*, 117(6), 3867-3878.

Dohen, M., Lœvenbruck, H., & Hill, H. (2006). Visual correlates of prosodic contrastive focus in French: Description and inter-speaker variabilities. In R. Hoffmann & H. Mixdorff (Eds.), *Proceedings of Speech Prosody 2006* (pp. 221-4). Dresden: TUD Press.

Perceptual Frequency and Formant Frequency in R speech

Naomi Nagy

Department of Linguistics, University of Toronto, Canada; nn.in.nh@gmail.com

Based on auditory analysis of 10,000 tokens from over 50 speakers, Irwin & Nagy (2007) established that postvocalic (R) is undergoing a change in the Boston area: younger and more educated speakers have higher rates of retention of post-vocalic [ɹ] in words like "park" and "car". As with most North American sociolinguistic studies of consonantal variation, this investigation began with binary auditory judgments of the presence or absence of [ɹ]. However, given that [ɹ] alternates with a vocalic allophone, and given that it is the most vocalic of the consonants, acoustic analysis and consideration of (R) as a scalar variable should be fruitful and contributes to our understanding of phonetic cues to morphological structure and social information. This approach also permits a better understanding of the transmission of variable patterns within a community, addressing a particular "Poverty of the Stimulus" problem related to the acquisition of variable patterns: correlations are found between the acoustic frequency of [ɹ]s produced in varying (linguistic and social) contexts and the frequency with which /ɹ/s are heard as [ɹ] (vs. Ø or a vocalic allophone) in these contexts. Thus speakers receive two different types of input to determine the appropriate frequency of use of each variant in different linguistic and social contexts.

My illustration of this advance is based on comparisons between two types of frequencies. The first, which I refer to as *Perceptual Frequency*, is the rate of perceived occurrence of [ɹ], or how often listeners (coders) hear [ɹ] for (R). Perceptual Frequency rates were calculated for various lexical items, for linguistic variables related to phonological context and morphological position, and for several social variables: age, sex, ethnicity, education and occupation. The relationships between Perceptual Frequency and these variables are examined via regression analysis.

The second type of frequency, *Formant Frequency*, is explored in several ways. Consonantal [ɹ] production is frequently associated with a lowered F3 in the literature. However, this measure does not produce a clear distinction, in these data, between tokens heard by the coders as r-ful vs. those heard as r-less, nor does it allow for controlled comparison across speakers. Thus several frequency measures were made: F3, F3-F2, F2/F3, as well as measures taking into account the duration of each segment and the F2 and F3 of onset (non-variable) /ɹ/. Correspondences between these measures and the perceived presence/absence of [ɹ will be presented.

I explore the relationship between the two types of frequency, building on Hay & Maclagan's (*forthcoming*) finding, in their study of intrusive [ɹ] in New Zealand English, that there is greater constriction (lower F3) in contexts that favor [ɹ] more often. They account for this in the context of Exemplar Theory, which predicts lower Formant Frequencies in tokens that occur in contexts associated with higher rates of Perceptual Frequency. In other words, the "strength" of [ɹ] can be meaningfully interpreted in two related ways: how often [ɹ] is heard in a particular context (Perceptual Frequency) and how constricted the [ɹ] is in each production (Formant Frequency).

This provides valuable insight into how speakers might acquire accurate information about the frequency with which a certain type of token should be produced in an r-ful vs. r-less manner, whether in terms of lexical, social, or linguistic context, without hearing a huge number of examples. The lower the F3 value of a particular r-ful token, the more likely that type is to be produced with an [ɹ] (vs. an unconstricted variant). That is, speakers may acquire Perceptual Frequency knowledge as a byproduct of Formant Frequency knowledge, resolving the Poverty of the Stimulus problem.

To illustrate, Table 1 gives the correlation scores (r^2) of Formant Frequency (the average minimum F3 value in the tokens perceived to be r-ful) to Perceptual Frequency (% tokens perceived as r-ful) for different breakdowns of the sample. These correlations show that listeners receive correlated cues to the "appropriate" rate of [ɹ]–production in different contexts as they hear different speakers and linguistic contexts: contexts where [ɹ] appears more frequently have acoustically distinct forms of [ɹ].

Table 1: Correlations between Formant Frequency and Perceptual Frequency

Context	r^2
Individual speakers (most to least frequent deleters)	-0.32
Lexical items (least to most r-ful words)	-0.64
F3 of onset /ɹ/* (lowest to highest onset F3)	0.29

*This correlation suggests a perception effect on the part of the coders (cf. Stuart-Smith 2007).

Correlations for individuals grouped by various social categories will also be reported. Further studies may show whether this type of correlation between Formant Frequency (or other scalar measures of production or perception) and Perceptual Frequency exists for other variables. The careful acoustic analyses of alveolar obstruents conducted by Foulkes, Docherty, et al., and North American vowel analyses in the Labovian tradition would be prime candidates.

A note on method

56 speakers from Boston and southern New Hampshire were digitally recorded reading a light-hearted story containing 224 words with post-vocalic /ɹ/. Environments were categorized according to morphological position, phonological context, and the interacting factors of word class (functional/lexical), word length, and lexical frequency. Two coders listened to each token a number of times and coded it as r-ful or r-less, checking each others' work and reaching an intercoder reliability rate of ≥85%. The factors were submitted to multivariate logistic regression analysis. All linguistic factors except word class proved significant, with perceived [ɹ] vs. Ø as the independent variable.

We acoustically analyzed a subset of tokens. Data for 20 of these speakers (10 White, 10 African-American, all Bostonian) and 20 words (all of similar lexical frequency and all containing /iɹ/ or /aɹ/ in a stressed syllable closed by another coronal consonant, *e.g.*, "fierce", "card") were selected. F2 and F3 and the duration of the /iɹ/ were measured. Linear regression analyses were conducted to examine the relationships between these continuous acoustic measurements and the independent variables above. Similar patterns of significant effects were found in the binary and scalar approaches.

References

Docherty, G.J., & Foulkes, P. (2005). Glottal variants of (t) in the Tyneside variety of English: an acoustic profiling study. In W. Hardcastle & J. Mackenzie Beck (Eds.), *A Figure of Speech – a Festschrift for John Laver* (pp. 173-199). London: Lawrence Erlbaum.

Docherty, G.J., & Foulkes, P. (2001). Variability in (r) production – instrumental perspectives. In H. Van de Velde & R. van Hout (Eds.), *'r-atics: Sociolinguistic, Phonetic and Phonological Characteristics of /r/* (pp. 173-184). Brussels: ILVP.

Foulkes, P., & Docherty, G.J. (2006). The social life of phonetics and phonology. *Journal of Phonetics*, 34(4), 409-438.

Hay, J., & Maclagan, M. (*forthcoming*). Social and phonetic conditioners on the frequency and degree of 'intrusive /r/' in New Zealand English. In D. Preston & N. Niedzielski (Eds.), *Methods in Sociophonetics*..New York: Mouton de Gruyter.

Irwin, P., & Nagy, N. (2007). Bostonians /r/ speaking: A quantitative look at (R) in Boston. *Penn Working Papers in Linguistics 13: Selected papers from NWAV 35,* 135-147.

Stuart-Smith, J. (2007). *A sociophonetic investigation of postvocalic /r/ in Glaswegian adolescents*. Paper presented at NWAV 36.

When does variation lead to change? A dynamical systems model of a stress shift in English

Partha Niyogi* & Morgan Sonderegger#

*Departments of Computer Science and Statistics, University of Chicago, USA; niyogi@cs.uchicago.edu
Department of Computer Science, University of Chicago, USA; morgan@cs.uchicago.edu

Introduction

All language change begins with variation, but most variation does not lead to change. In fact significant phonetic variation occurs even within the speech of single speakers (Pierrehumbert, 2003), suggesting a restatement of the "actuation problem": why does language change happen at all, why does it arise from variation, and what determines whether a pattern of variation between several forms is stable or unstable (leads to change)? To address these questions we present a case study of prosodic change in English between multiple forms, and model its dynamics using dynamical systems models for a population of learners (Niyogi, 2006). We propose that the distinction between unstable and stable variation corresponds to the distinction between models which show bifurcations (or "phase transitions") as model parameters change, and those which do not. These classes correspond to models with or without "ambiguity", meaning that learners receive some examples which are not clearly heard as one form or another.

Data

The English lexicon contains approximately 1000–2000 bisyllabic words with both noun and verb forms (contest, contrast, cement). The stress for a N/V pair is denoted (N, V) here (e.g. (1, 1) means doubly- initial stress). There are four possibilities for each pair: (1,1), (1,2), (2,2), or (2,1). Sherman (1975) found that approximately 150 N/V pairs have shifted stress since 1600, many to (1,2). Combining Sherman's data (1600-1800) with our own (1800-2005), we recorded the stress of these 150 words in 85 British and American dictionaries from 1600-2005. These sources often report variation. Considering the trajectories formed by the moving average of each word's reported stress, we found:

1. (2,1) is never attested.
2. The changes (2, 2) → (1, 2), (1, 1) → (1, 2), (1, 2) → (2, 2), (1, 2) → (1, 1) occur (in decreasing order of frequency), and (1, 1) ↔ (2, 2) do not occur.
3. Synchronic variation (for a particular N/V pair) is widespread, but long-term variation is rare.
4. Classes of words sharing a prefix (such as con-) have more similar trajectories than random classes (analogy), but only for larger prefix classes

We also found

5. that (1, 1), (1, 2), and (2, 2) are stable states: in a random subset of 100 of all N/V pairs, most have shown little synchronic variation, and no change, over 400 years.

Many linguistic explanations for (1)-(5) are possible, but testing whether a proposed theory diachronically explains (1)-(5) involves building formal models of the interplay between acquisition, variation, and change.

Models

In order to reason about the possible trajectories of language change in a population, we constructed a variety of models; here we focus on two classes of models. Both assume a probabilistic characterization of the internal variation on the part of speakers, and a population of learners each of whom learns from the population at large during a learning period. We use a probabilistic characterization based on evidence from a corpus of radio speech showing speaker-level variation in stress placement for N/V pairs.

Models with Phonological Ambiguity

Assume a word has two forms (say 1 and 2, the two stress patterns). Let each mature speaker have a value $x \in [0, 1]$ that characterizes the probability with which they produce form 1. Let the distribution of x values in the adult population be given by a probability distribution $P(x)$. Learning is the mechanism by which this variation is transmitted to the next generation. A learner situated in this population hears k utterances of the word from the population at large. On the basis of his/her experience, the learner develops a value of x which is used as a mature speaker.

Ambiguity, for us, means the following. If a speaker intended form 1, then with probability a, the hearer perceives form 1 but with probability $(1 - a)$, the hearer is unable to determine whether form 1 or 2 was intended. Similarly, form 2 is heard correctly with probability $(1 - b)$, and is heard as ambiguous with probability b. Then, if the learner hears form i k_i times and ambiguous forms $k - k_1 - k_2$ times, a natural probability matching algorithm on the part of the learner would be to use an internal variation x given by $x = k_1 / (k_1 + k_2)$ (provided $k_1 + k_2 > 0$; say x=0.5 otherwise). Then, if u_t is the average value of x in the generation t, one can calculate u_{t+1}, the average value in the next generation. In this case, for large k we have

$$u_{t+1} = \frac{au_t}{au_t + b(1 - u_t)}$$

This dynamics is (i) nonlinear (ii) has stable points where the population uses either form 1 or form 2 exclusively (iii) bifurcations between the two stable points as a function of $(a - b)$. We elaborate on this basic insight by considering the effects of (a) coupling between words (analogy between words during learning) (b) finite k (word frequency) (c) finite population size (d) mixed misperception and ambiguity.

Models without Phonological Ambiguity

In contrast, one might consider models without ambiguity. The setup is exactly as before except learners always interpret the acoustic stimulus categorically. In other words, if the speaker intended form 1, the learner will either hear it as form 1 or 2: there is no “ambiguous” state. Thus there may be misperceptions but none of the primary linguistic data is actually thrown out during the learning period. We consider models with this character and show that the dynamics that arise have very different properties. In particular, there is always one stable attractor that corresponds to a mixed state, i.e., stable variation persists. Further there are no bifurcations.

Conclusions and Contributions

Our main contribution is to collect a corpus of empirical data on phonological change over time, to expose the important aspects of language change in this corpus, to construct a variety of models of language acquisition, variation, and change to understand and reason about the trends we see in this corpus. Our main result is that if learners perceived their input with the possibility of phonological ambiguity, then the dynamics is nonlinear, there are bifurcations (“phase transitions”) that may explain the actuation problem, and the stable modes of the population are close to homogeneous. In contrast, if learners perceived their input in a purely categorical way, this is not true and in particular stable variation is possible. There are many subtle aspects of the dynamics of linguistic populations that we wish to expose through this kind of a case study of phonetic variation and phonological change.

References

Niyogi, P. (2006). *The Computational Nature of Language Learning and Evolution*. Cambridge: MIT Press.

Pierrehumbert, J. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech*, 46, 115-154.

Sherman, D. (1975). Noun-verb stress alternation: An example of the lexical diffusion of sound change in English. *Linguistics*, 159, 43–71.

Does what listeners hear affect what they say? Evidence from shadowing experiment

Naomi Ogasawara

Department of English, National Taiwan Normal University; naomi703@ntnu.edu.tw;

The relationship between speech perception and production has been accounted for in various ways. The Motor Theory (Liberman & Mattingly 1985) claims that what listeners perceive is the speaker's intended phonetic gestures rather than acoustic signals themselves. Speech perception is achieved by decoding acoustic signals to the articulatory movements. Motor Theory can well handle the problem of the lack of invariance. In psycholinguistics, phonological accounts (van Alphen & McQueen, 2006, among others) speculate that phonological knowledge interferes with the listeners' processing of speech sounds. Based on the phonological account, low-level fine phonetic details do not matter because what listeners process is phonemes rather than phones.

In this paper, we investigate the relationship between speech perception and production in relation to phonological interference. We address the following three questions in the study: (1) whether or not listeners perceive low-level allophonic details; and (2) If so, how a listeners' knowledge of phonology which restricts the conditions for allophonic variation, interferes with production when listeners repeat the utterance. To address these questions we investigated devoiced and voiced variants of the high vowel /i/ in Japanese. As is well known, vowel devoicing is a common phonological phenomenon in Japanese. Vowel devoicing occurs when high vowels (/i/ and /u/) are either between voiceless consonants or precede a voiceless consonant (e.g., [ki̥ta] 'north') and followed by a pause at the word-final position (e.g., [desu̥] COPULA-be). On the other hand, when vowels are adjacent to at least one voiced consonant, vowel devoicing does not happen (e.g., [mita] 'saw' or [deru] 'out').

In this study, a shadowing task together with auditory priming was employed. Thirty-four native speakers of the Tokyo dialect were recruited in the Tokyo area. They were asked to listen to thirty-six lexical word pairs with devoiced and voiced vowels in the devoicing and voicing environments (e.g., /akikaN/ 'empty can' for devoicing environment and /itʃigo/ 'strawberry' for the voicing environment). Acoustically all devoiced vowels in the stimuli were the deleted kind, which had neither periodic waveform nor vertical striations in the spectrogram. Only the coarticulation cues remained, mainly in the preceding consonant. The stimuli were made by splicing the target mora (e.g., [kʲ] and [ki]) so that segments other than the target mora were identical between devoiced and voiced stimuli. The first thirty-six stimuli served as primes and the other thirty-six stimuli served as targets. There were three priming conditions (match: exactly the same stimulus was presented as a prime and a target; mismatch: the stimuli with different vowel realizations (devoiced and voiced) were presented as a prime and as a target; control: an unrelated word was presented as a prime and an experimental stimuli was presented as a target). The participants were asked to say what they heard in terms of their usual pronunciation as soon as they recognized the word, and their speech was recorded.

Their productions of the target vowels were analyzed. The results show that in the devoicing environment, shadowers produced devoiced vowels (phonologically appropriate) about 86 % of the time after they heard devoiced stimuli. Shadowers were still able to produce devoiced vowels far above chance (76 % of the time) even after they heard voiced stimuli which were phonologically inappropriate. However, perception of the inappropriate stimuli significantly dropped the rate of the production of the appropriate vowel ($F(1, 27) = 5.933$, $p < 0.05$; $F(1, 27) = 7.399$, $p <0 .02$ for shadowing of prime stimuli and of target stimuli, respectively). Also, in the same environment, shadowers produced inappropriate (voiced) vowels significantly more when they heard the inappropriate stimuli than when they heard the appropriate stimuli ($F(1, 27) = 6.239$, $p <0 .02$; $F(1, 27) = 11.164$, $p < 0.005$ for shadowing of prime stimuli and of target stimuli, respectively). However, in the voicing environment, hearing inappropriate (devoiced) stimuli did not affect the production of the appropriate vowel. These results were seen when repeating both primes and targets. Moreover, statistical analysis with the target shadowing revealed that there was no significant effect of priming conditions ($F(2, 54) = 8.73$, $p > 0.1$). What affected production of vowels was what shadowers heard right before their production rather than what they heard a while ago (in the priming block). This was surprising, but it may be because the materials were familiar words for the subjects, thus, the pronunciation of the words was consistent regardless whether or not they heard the words before.

The above results posit that listeners are quite sensitive to low-level allophonic differences and that the phonological appropriateness of an allophone in the heard stimuli has a significant effect on the appropriate production of said allophone. This interpretation is disfavored within the Motor Theorist account because the theory predicts that what participants hear should equally affect what they say regardless of the conditions for the allophonic variation. However, this was not the case. Also that allophonic variation is quite important both for perception and production disfavors the phonological account in which the importance of phoneme not phones is emphasized for word recognition. Moreover, the results indicate that vowel devoicing in Japanese has not become a very solid (still gradient) phonological phenomenon in Japanese because shadowers still produced (inappropriate) voiced vowels in the devoicing environment, whereas this was not the case in the voicing environment. In the voicing environment, inappropriate (devoiced) stimuli did not affect production of the appropriate vowel. This may be because voiced vowels are underlying forms and devoiced vowels are not. Together with these results, in psychological reality, listeners are able to hear phonetic details, which affect production of non-underlying form. This may suggest that the strength of the link between perception and production varies depending on whether the target sound is underlying or non-underlying.

References

Liberman, A. M., & Mattingly, I. (1985). The motor theory revised. *Cognition*, 21, 1-36.

Von Alphen, P. M., & McQueen, J. M. (2006). The effect of voice onset time differences on lexical access in Dutch. *Journal of Experimental Psychology: Human Perception and Performance*, 32, 178-196.

Lexical Tonal Space and Sandhi Rule

Ho-hsien Pan*, Pi-chiang Li#, Chia-Chien Lee*, Hsiao-Han Huang* & Hui-ying Tsou*

* Department of Foreign Languages and Literatures, National Chiao Tung University, Taiwan; hhpan@faculty.nctu.edu.tw
# Department of Statistics, Feng Chia University, Taiwan; lipc@fcu.edu.tw

The access of lexical meaning in tone languages requires identification of lexical tones. Gandour (1978) in a Multidimensional Scaling (MDS) study found that in addition to f0 cues, tone sandhi also influenced tone perception in Cantonese, Mandarin and Taiwanese. This is because application of sandhi rules between level and contour tones can lead to perceptual confusion between related tones. Later MDS studies on tonal perception in three varieties of Mandarin, namely Beijing, Rugao, and Yantai, also revealed perceptual proximity for lexical tones linked by sandhi rules (Johnson, 2004; Huang, 2004). Following the assumption that tones linked by sandhi rules are closer in perceptual distance, it is proposed that Taiwanese lexical tones would be perceptually similar as well due to chain sandhi rules, HH and LH -> MM -> ML -> HL -> HH. That is, the HH and LH tones become more like the MM tone which in turn becomes more like the ML tone. The ML tone is close to both the MM and HL tones, whereas the HL tone is close to the ML and HH tones which are close to the MM tone. Hsieh (1976) questioned the psychological reality of the Taiwanese sandhi chain and suggested that by memorizing two independent lexical entities that were analogically associated with the same morpheme, Taiwanese speakers did not access the juncture and non-juncture lexical forms through phonological processes but through allomorphic process. For example, the morpheme /kun/ "near" is produced with a MM tone when produced at a phrase final position and with a ML tone at a phrase initial or medial position in a tone group. Hsieh suggests that these two surface realizations are perceived as allomorphs of a single morpheme. This study intends to explore the psychological reality of the Taiwanese sandhi chain. If there is a sandhi chain in Taiwanese, then a circular tonal distribution should be observed in the perceptual tonal space. If it is not, then the Taiwanese sandhi chain is under question. Although Gandour (1978) studied how speakers of English, a stress language, perceived lexical tones, it was unknown how speakers of pitch accent languages, e.g. Swedish, perceived lexical tones. By comparing tonal spaces between Taiwanese and Mandarin speakers, the influences of different sandhi backgrounds on tonal space can be seen, whereas the comparison between data from Swedish and tone languages speakers would show the influence of auditory perceptual mode on tone spaces.

Fourteen Taiwanese, 10 Mandarin, and 13 Swedish subjects participated in tonal comparison experiments. Subjects clicked "same" or "different" buttons after comparing two syllables of /kun/ carrying either two juncture, sandhi, citation tones, or one juncture tones versus one sandhi tone. In addition, resynthesized pitch only stimuli were also compared in an AX experimental paradigm. Using MDS analysis, the perceptual tonal space was derived from the reciprocals of response times for true responses.

Table 1: Natures of x-, y-axis and presence of tone circles in MDS plots. T: Taiwanese, M: Mandarin, S: Swedish. Direction: falling vs. non-falling tones

T	F0 onset	F0 range	Yes	F0 offset	F0 onset	Yes	Direction	F0 onset	Yes
M			No	Direction	F0 onset	No	Direction	F0 onset	Yes
S	F0 onset		Yes	Direction		No		F0 onset	Yes

As shown in Table 1, Taiwanese speakers used f0 onset, f0 offset, f0 range, and the distinction between falling and non-falling tones as perceptual cues; whereas Mandarin and Swedish speakers used only f0 onset and the distinction between falling and non-falling tones to differentiate lexical tones. As shown in Table 1, although a circular tonal distribution was consistently observed in the Taiwanese speakers' data across all sessions, in the Mandarin speakers' data, circular tonal distributions were found only among sandhi tones (Figure 1). By placing tones of the same contours far away from each other, Mandarin speakers treated juncture and sandhi tones of similar contours as two different entities. As for the Swedish speakers' data, although a circular tonal distribution was found in juncture, sandhi, and pitch-only sessions, Swedish speakers' error rate was 14.8%, compared with 7.4 % for Mandarin speakers and 5.3 % for Taiwanese speakers.

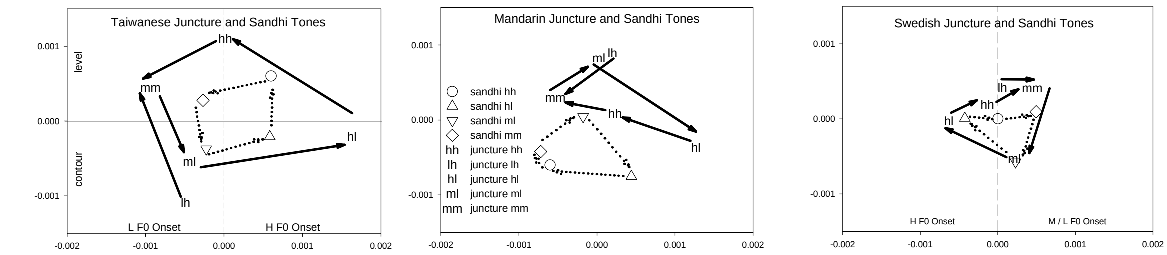


Figure 1: Lexical tonal space for Taiwanese, Mandarin and Swedish subjects.

These results suggest that speakers are influenced by their native languages. Since the sandhi rule HH -> MM linked two level tones, the MM ->ML rule linked two tones with medium f0 onset, the ML -> HL linked two falling tones, and the HL -> HH linked two tones with high f0 onset, Swedish speakers experienced with lexical accent use pitch as an extra-linguistic cue to form tone circles and cluster tones of same contours together among the tonal pairs receiving true responses. Mandarin speakers experienced with lexical tone separate corresponding juncture and sandhi tones, e.g. HH juncture and sandhi tones, into different tonal space suggesting a lack of phonological linkage. Unlike Mandarin speakers, Taiwanese speakers form tonal circles across all sessions and cluster tones of same f0 contour together. Moreover, the larger tonal circles show the clearer perceptual tonal distinction by Taiwanese speakers than by Mandarin or Swedish speakers. In sum, (1) Taiwanese sandhi rules are auditory based. Even Swedish speakers can form tone circles. However the larger number of errors and the absence of a tonal circle in the citation session imply the lack of a robust phonological linkage between sandhi and juncture tones for Swedish speakers. (2) Being experienced with lexical tones, but inexperienced with Taiwanese sandhi rules, Mandarin speakers perceive corresponding sandhi and juncture tones as separate entities and form a tone circle for only sandhi tones, (3) The robust presence of large tone circles across all sessions reveal the psychological reality of sandhi chains for Taiwanese speakers.

Acknowledgements

This research was supported by National Science Council, Taiwan (NSC94-2411-H-009-011).

References

Gandour, J. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics*, 11, 149-175.

Gandour, J. (1978). Crosslanguage differences in tone perception: A multidimensional scaling investigation, *Language and Speech*, 21(1), 1-33.

Huang, T. (2004). *Language Specificity in Auditory Perception of Chinese Tones*. Ph.D. dissertation, The Ohio State University.

Hsieh, H.-I. (1976). On the reality of some phonological rules. *Lingua*, 38, 1-19.

Johnson, K. (2002) *Acoustic and Auditory Phonetics*. Oxford: Blackwell.

Intrinsic Pitch is not a universal phenomenon: Evidence from Romance languages

Daniel Pape* & Christine Mooshammer#

*Centre for General Linguistics (ZAS), Berlin, Germany; pape@zas.gwz-berlin.de
#Haskins Laboratories, New Haven, USA; tine.mooshammer@hasikins.yale.edu

This work examines the relation and interaction between speech production and speech perception with respect to the micro-prosodic phenomena “Intrinsic F0” versus “Intrinsic Pitch”.With regard to speech production, Intrinsic F0 describes the dependence of the fundamental frequency on vowel height: Other things being equal, close vowels show a higher F0 compared to open vowels. With regard to speech perception, Intrinsic Pitch describes the dependence of the perceived vowel pitch on the height of the presented vowel: Close vowels are perceived with a lower pitch impression compared to open vowels. The aim of this work is to investigate similarities and differences by conducting a combined examination of the phenomenon “IF0-IP”, comparing cross-linguistic influences on the other hand and musical education influences on the other hand.

Research on speech production showed that Intrinsic F0 differences are found throughout all languages and language families (Whalen and Levitt 1995): Stress-timed languages, syllable-timed languages and tone languages show nearly the same Intrinsic F0 amount and distribution. This finding underpins a biomechanical explanation on basis of the human vocal tract properties, although other theories argue for an active increase of acoustic differences between different vowels.

Research on speech perception generally acknowledges that perceptual Intrinsic Pitch differences are found for Germanic languages: Due to the strong evidence in speech production data, it is argued (Fowler and Brown 1997) that listeners compensate for automatically occurring F0 height differences, which might lead to difficulties for stable prosodic parsing. However, other explanations for Intrinsic Pitch claim that the pitch differences are introduced due to the special properties of the human auditory system (psychoacoustic theory), e.g. masking and non-linearity. Silverman (1987) showed that in phrasal context Intrinsic Pitch can be used to evoke accent differences perception by changing the IF0 height of stressed vowels.

Interestingly, empirical Intrinsic Pitch data is only available for the Germanic languages English and German, although it is generally claimed to be a universal phenomena. In the current study we aim at a cross-linguistic validation of Fowler´s compensation theory by expanding experimental data on Intrinsic Pitch to typologically diverse languages such as Romance and tone languages. Previous studies (Pape et al. 2005) gave evidence for a strong cross-linguistic difference when comparing Intrinsic Pitch results for the Romance language Catalan with German. Furthermore, the before-mentioned psychoacoustic theory could not be supported by our data.

Apart from the linguistic background of the listener the phenomenon of Intrinsic Pitch differences has been found to be strongly affected by musical education (Rauscher and Hinton 2003). Therefore this experiment was carried out in a fully crossed experimental design with the factors language background and musical education (professional musicians versus non-musicians). We used a standard 2I2AFC procedure, testing for two different isolated vowels in all possible orders in a single vowel pair comparison. The first experimental part tests for the German vowel pair /iː ɑː/ which maximally differs in vowel height. The second part examines vowel pairs for the listeners’ native language (thus e.g. Spanish vowels /iː ɑː/ for Spanish subjects). The F0 differences to be tested were PSOLA pitch-shifted in the range of ±10Hz (2.5Hz steps), proven to be sufficient to drive the full amount of Intrinsic Pitch differences. The terminology was absolutely clear for the participants in their native language and the Romance language listeners were not educated in one of the Germanic languages. For each condition around 30 listeners were tested. The experiment used isolated vowels (cut from a stressed /mVme/ context), since only so a consistent interpretation without inclusion of macroprosodic phenomena could be guaranteed. Clearly, the continuation of Silverman’s experiments is promising, but due to the necessary inclusion of a large parameter space is not undertaken in the present experimental series.

The main new finding is that the number of listeners not sensitive to the given pitch task is much higher for all Romance languages than for German. Spanish and Portuguese have the highest number with 46% of non-musicians showing *pitch-insensitivity*. Italian with 23% is in between, and German shows the lowest number with 6%. Thus, here for nearly half of the Romance language listeners Intrinsic Pitch differences cannot be extracted. This gives clear evidence that Intrinsic Pitch is a language-dependent phenomenon. The insensitivity found is independent of the stimulus identity, which could be shown in the comparison of the German and native stimuli conditions. Thus, the native stimulus condition did not increase sensitivity.

When analyzing the valid Intrinsic Pitch values for the *remaining* pitch-sensitive listeners, the following results are obtained: When pooled over all German listeners, for the vowel height comparison /i ɑ/ it is found that the values significantly differs, although with a lower value than in the classical literature. However, when separately examining the different populations, it is seen that no statistical difference is obtained, neither for the professional musicians nor non-musicians. Thus, Intrinsic Pitch is generally present in German, but the significance of the phenomenon can not be shown when examining populations contrasting in the (important) factor musical education.

With regard to cross-linguistic differences, it is found that for all languages Intrinsic Pitch was absent for professional musicians. Further, when examining the occurring differences in musical education, it could be shown that across all languages the values significantly differ when comparing professional musicians with non-musicians. Thus, here important evidence is given that increased musical education significantly influences the occurrence and distribution of Intrinsic Pitch. For the non-musicians, only inconsistent patterns are found, For Italian the results are clearer: Here, Intrinsic Pitch is present for the native vowel condition. Thus, Italian presents a special case: It seems that Intrinsic Pitch is present only for the population of non-musicians and the native vowels.

In summary, it is given evidence that Intrinsic Pitch definitely has to be classified in a language-specific manner: For the Romance languages Spanish and Portuguese, Intrinsic Pitch is mostly not verifiable. This finding points to the idea that speakers of Germanic languages indeed use a cue to parse the vowel pitch into a prosodic and vowel height component, and therefore are completely able to solve the pitch tasks asked for in our experimental conditions. Our findings on German also corroborate the idea of micro-prosodic compensation is valid, probably with the aim of compensating for speech production F0 differences.

Since it is shown in the literature that Romance languages use F0 similarly for prosodic marking, our results seem to indicate the following hypothesis: In Romance languages, Intrinsic F0 as an additional vowel height cue is not discriminately used, resulting in the observed pitch-insensitivity for the Romance languages Spanish and Portuguese. Whether this lack of the cue Intrinsic Pitch is due to the much smaller vowel inventory in these languages or to other phenomena will be examined in further research. The intermediate position for the Italian language supports the ideas of Gili Fivela and Zmarich (2005): They showed by articulatory data that Italian is a mixture of stress-timed and syllable-timed languages. Our results confirm this typological classification with the Italian listeners showing a higher pitch-discrimination and more stable Intrinsic Pitch values compared to Spanish and Portuguese, but significantly worse than the German listeners.

References

Fowler C.A., & Brown J.M. (1997). Intrinsic f0 differences in spoken and sung vowels and their perception by listeners. *Perception and Psychophysics*, 59(5), 729-738.

Gili Fivela B., & Zmarich C. (2005). Italian geminate under speech rate and focalization changes: kinematic, acoustic and perception data. *Proceedings of INTERSPEECH’2005-EUROSPEECH*, 2897-2900, Lisboa, Portugal.

Pape D., Mooshammer C., Fuchs S., & Hoole, P. (2005). Intrinsic pitch differences between German vowels /i:/, /I/ and /y:/ in a cross-linguistic perception experiment. *Proceedings of the ISCA workshop on plasticity in speech perception* (PSP2005), London, UK

Rauscher H., & Hinton S.C. (2003). Type of music training selectively influences perceptual processing. *Proceedings of the European Society for the Cognitive Sciences of Music*, Hannover, Germany

Silverman K.E.A. (1987). *The structure and processing of fundamental frequency contours*., PhD dissertation, University of Cambridge, UK.

Whalen D., & Levitt A. (1995). The universality of intrinsic F0 of vowels. *Journal of Phonetics,* 23, 349-366.

Frequency of use and the Apparent Time Paradigm

Nicolai Pharao

University Department of Nordic Studies and Linguistics, University of Copenhagen, Denmark;
nicolaip@hum.ku.dk

On-going sound change has almost exclusively been studied through the Apparent Time Paradigm, in which sound change is charted by analyzing the phonetic variation of speakers of different age groups and inferring from such differences the course of an on-going change in the realization of particular phonemes. Most of these changes have been studied at the level of individual segments, e.g. /t,d/ deletion, /a/ raising, and /u/ fronting, regardless of which word they occurred in, while still factoring in phonological context and phonotactic or morphological constraints (e.g. position in the syllable or position in stem or affix; cf. Raymond, Dautricourt & Hume, 2006). In addition to this, some phonetic variation has been observed to be specific for a sub-group of all the words containing the conditioning environment, a mechanism known as Lexical Diffusion (cf. Wang, 1977; Yaeger-Dror, 1994).

In a usage based approach to sound change, e.g., (Bybee, 2001), most sound changes are seen to be gradual in the sense that they do not apply across the board to every word containing the conditioning environment, but start in high frequency words and proceed gradually through the lexicon; i.e., most sound changes are diffused lexically. This hypothesis has implications for the study of sound change through comparison of samples of speakers of different age groups. An allophone that is hypothesized to be participating in an on-going change should be attested only in words of relatively high frequency in the older generation as compared to its pattern of distribution in a sample taken from younger generations in the same speech community. That is, the new allophone should be attested in a broader range of words in the younger generation than in the older generation. Furthermore, if frequency of use is operative in the diffusion of a sound change across generations, we can also ask about the role of frequency in the diffusion of a sound change in individual lexicons over time. If individual lexicons are affected by frequency of use in the manner outlined in (Bybee, 2001), the innovative allophone should become relatively more frequent in items of lower frequency later in life.

The present study investigates the role of frequency in the phonologies of individuals and groups through the use of a sub-corpus of a Real Time study currently being conducted of spoken Danish. Real Time studies involve re-recording and re-analysis of the behavior of subjects from an original Apparent Time sample (see Sankoff, 2006, for a review of the literature on Real Time studies.) This enables an investigation of the stability of phonetic behavior across the lifespan. The sub-corpus contains auditory analyses of the variation in the realizations of /ɛ/, /æ/ and /ɑ/ in different phonetic contexts in the speech of 5 different subjects who were originally interviewed in the 1980s and have been re-interviewed in 2005. The interviews, ranging from 30 minutes to 2 hours in length, have been analyzed in their entirety, including classification of every stressed realization of the phonemes. This enables a test of the role of frequency at the time of original recording through the comparison of the range of frequencies of words containing standard and innovative allophones of the phonemes hypothesized to be changing. The role of frequency in possible lifespan changes can also be studied by investigating whether incipient sound changes at the time of the original recording have spread only to words of lower frequency in the speech of individual speakers.

Should individual speakers prove to have adopted innovative allophones in lexical items of lower frequency later in their lives, this would lend support for the role of frequency in the continual shaping of the individual lexicon. If the occurrence of the innovative allophones do not spread to words of lower frequency in the re-recordings of the informants, this would suggest that frequency and hence the mechanism of lexical diffusion is operative at the level of groups of speakers only, and that, while the phonological grammars of individuals may be gradient, lexical conditions on the distribution of phonetic innovations would seem to fossilize some time before adulthood.

References

Bybee, J. (2001). *Phonology and Language Use*. Cambridge: Cambridge University Press.

Raymond, W.D., R. Dautricourt, & E. Hume (2006). Word-internal /t,d/ deletion in spontaneous speech: Modeling the effects of extra-linguistic, lexical, and phonological factors. *Language Variation and Change*, 18 (1), 55-98.

Sankoff, G. (2006). Age: Apparent time and real time. In K. Brown (Ed.), *Elsevier Encyclopedia of Language and Linguistics*, 2nd edition, (Article Number: LALI 01479). Amsterdam: Elsevier.

Wang, W. S-Y. (1977). *The Lexicon in Phonological Change*. Berlin: Mouton.

Yaeger-Dror, M. (1994). Phonetic evidence for sound change in Quebec French. In P. Keating (Ed.), *Phonological Structure and Phonetic Form – Papers in Laboratory Phonology III* (pp. 267-292). Cambridge: Cambridge University Press.

The Scope of Phrasal Lengthening in Articulation: Prosody and Prominence

Daylen Riggs and Dani Byrd

University of Southern California

Introduction

Over the preceding Laboratory Phonology conferences, experimental work has clearly established that abstract phonological structure is directly reflected in the spatiotemporal details of speech production. In fact this softening of the clear severance between phonetics and phonology and an appreciation of the complexity of their relationship has been a prominent contribution of the LabPhon tradition. The study of the phonetics-phonology interface has been extended to the phonetics-prosody interface, and we have, again, seen that abstract structural properties, this time of utterances rather than words, have a complex and rich effect on the articulatory details of speech production (e.g., Cho, 2006). This has led to new theoretical conceptions regarding how phrase boundaries should be represented and to new computational models of their realization in speech (e.g., Byrd & Saltzman, 2003). In the study described below, we consider how two related but independent aspects of prosodic structure—prominence and boundaries—interact in shaping articulatory timing.

The lengthening of phonological units at phrase edges has long been recognized in the acoustic domain and has also been observed in the articulatory domain (e.g. Byrd and Saltzman, 1998; Cho, 2006). However, few studies have examined at what distance from the phrase edge and under what conditions the effects of phrasal lengthening occur—the scope of lengthening. (See Turk and Shattuck-Hufnagel, 2007 for a review of the acoustic findings; little articulatory data exists except Krivokapic, 2007). Lexical stress and/or pitch accent are known to interact with acoustic phrasal lengthening. Shattuck-Hufnagel and Turk (1998) found that acoustic lengthening could extend leftward (pre-boundary lengthening) to reach the most prominent syllable in a boundary-adjacent word. Similarly, Turk (1998) found that unstressed-boundary adjacent syllables do not undergo acoustic phrasal lengthening but that stressed syllables near (but not adjacent to) a boundary do. The articulatory movement tracking study described below addresses the scope of phrasal lengthening in English as a function of the position of pitch accent adjacent to and at small distances before and after an intonational phrase boundary. In addition to examining the behavior of the pitch-accented syllable, it is designed to further investigate whether prosodic lengthening is evidenced for constriction gestures intervening between the phrase boundary and the nearby pitch-accented syllable.

The prosodic (or π-) gesture model is a theoretical account of spatiotemporal effects on articulation in the neighborhood of prosodic boundaries that has been developed by Byrd, Saltzman and colleagues in Lab Phon meetings V, VIII, and X. In the π-gesture framework, Articulatory Phonology's conception of phonological units as inherently dynamic is extended to the prosodic level in that prosodic boundaries are viewed as having a temporal extent or interval of activation during which the π-gesture acts to slow the pacing of co-occurring constriction activation trajectories. If the π-gesture is fixed in its temporal anchoring to the boundary and in its temporal field of activation, remote effects sensitive to pitch accent should not be observed. If, however, the coordination of the π-gesture with constriction gestures is malleable (see e.g., Byrd and Saltzman 2003), pitch accent in a syllable nearby a phrase boundary might attract the π-gesture—effectively re-coordinating it—or it might change the temporal extent of the π-gesture by having its pre- (or post-)boundary edge "reach-out" in time to the pitch-accented syllable—starting earlier or ending later. We call these possibilities, respectively, the shift account and the extension account. The study below examines these two possibilities based on the pattern of lengthening in both the pitch-accented syllable and in unaccented syllables intervening between the accented syllable and the phrase edge. Articulatory durations of gestures in these syllables are compared to the durations of gestures in identical pitch-accented syllables that appear in a control condition having no boundary. The extension account predicts boundary-related lengthening of gestures both in prominent syllables near a boundary and in syllables intervening between the boundary and the nearby prominent syllable. The shift or recoordination account predicts that while gestures in prominent syllables near a boundary may lengthen, gestures in intervening syllables will exhibit no or minimal lengthening.

An Articulatory Study of the Interaction of Phrase Boundaries and Prominence

We conducted an articulator movement tracking (EMA) experiment to examine both pre- and post-boundary contexts with pitch-accent placed 1, 2 and 3 syllables away from the boundary. Sentences with an intonational phrase boundary were paired with and compared to no-boundary control sentences of parallel phonological make-up and matching target words. The target syllables contained [p] and [m] gestures and the Euclidean lip aperture signal was created using the upper lip and lower lip *x-y* position signals. Constriction formation duration was defined as the duration between the velocity zero-crossings in the lip aperture signal for constriction onset and for constriction target achievement. Unpaired, one-tail t-tests were conducted to compare the means of test vs. control conditions for constriction formation duration. Three subjects participated.

The results indicate that when the pitch-accented syllable was immediately post-boundary, all subjects showed significant lengthening relative to the constriction duration in the matched pitch accented syllable in the no-boundary control condition. When the pitch-accented syllable was two syllables post-boundary, all subjects showed phrasal lengthening in the lip closing gesture contained in the syllable immediately adjacent to the boundary; however, no subjects showed lengthening in the pitch-accented syllable. When pitch-accent was placed three syllables post-boundary, two subjects showed phrasal lengthening in the gesture immediately adjacent to the boundary. One subject showed lengthening in the syllable intervening between the boundary and the pitch-accented syllable; none lengthened the gesture in the accented syllable at the 3-syllable distance relative to the no-boundary condition. The pre-boundary condition showed more remote boundary related lengthening than the post-boundary condition, at least for one subject. When pitch accent was placed two syllables before the boundary, one subject had boundary-driven lengthening of gestures in both the pitch-accented syllable and the boundary-adjacent syllable. When pitch-accent was placed three syllables pre-boundary, this same subject had lengthening of gestures in all three syllables—the pitch accented syllable, the intervening syllable, and the boundary-adjacent syllable. It is important to note that the magnitude of lengthening was greater for gestures in syllables closer to the boundary than those further away (see Krivokapic, 2007).

These patterns of results support the extension hypothesis of the π-gesture discussed above. The fact that boundary-related lengthening was always greatest in the immediately post-boundary syllable indicates that the π-gesture remains anchored or centered at the boundary. However, the fact that phrasal lengthening is sensitive to the placement of a nearby pitch accent and extends, at least for one of the subjects, leftward or rightward to become coproduced with prominent (pitch-accented) phonological material supports the π-gesture extension possibility. Boundary-adjacent lengthening effects were the most robust in the boundary adjacent syllable for this subject with the magnitude of lengthening less the further away from the boundary the gesture was located. This pattern of results is predicted by the π-gesture extension hypothesis, in that the π-gesture's activation grows as the phrase edge is approached and is strongest closest to the edge.

This study extends the on-going efforts of the LabPhon community to understand the relation between abstract linguistic structure and phonetic detail by examining the interaction of two different aspects of prosodic structure—prominence and boundaries—in shaping low-level articulatory timing. Simultaneously, it seeks to test and refine a theoretical model of the prosody-phonetics interface.

References

Byrd, D., & Saltzman, E. (1998). Intragestural dynamics of multiple phrasal boundaries. *Journal of Phonetics*, 26, 173-199.

Byrd, D., & Saltzman, E. (2003). The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening, *Journal of Phonetics*, 31,2, 149-180.

Cho, T. (2006). Manifestation of prosodic structure in articulatory variation: Evidence from lip kinematics in English, In L. M. Goldstein, D.H.Whalen, & C.T. Best (Eds.), *Laboratory Phonology 8: Varieties of Phonological Competence* (pp. 519-548). Berlin: Mouton de Gruyter.

Krivokapic, J. (2007). *The planning, production, and perception of prosodic structure*. USC dissertation.

Turk, A., & Shattuck-Hufnagel, S. (2007). Multiple targets of phrase-final lengthening in American English words, *Journal of Phonetics*, 35,445-472.

Obligatory release and stiffness modulation in Moroccan Arabic

Kevin Roon[1,2], Adamantios I. Gafos[1,2], Phil Hoole[3], & Chakir Zeroual[4,5]

[1]New York University, Department of Linguistics; [2]Haskins Laboratories, USA;
[3]IPS, Ludwig-Maxmilians-Universitaet, Muenchen, Germany;
[4]Univ. Sidi Mohamed Ben-Abdellah-Morocco; [5]Lab. de Phonétique et Phonologie, Univ. Paris III, France.
kdroon@nyu.edu

The present study provides evidence that the linguistic production system adjusts articulation timing online to achieve a categorical linguistic goal. Specifically, it presents data from Moroccan Arabic supporting the hypothesis that the dynamical parameter of stiffness (Munhall, Ostry, & Parush, 1985) is modulated during production to assure release in heterorganic stop-stop clusters in this language.

Dell & Elmedlaoui (2002: 231) describe a required "audible release" between two heterorganic stops in a cluster in Moroccan Arabic. In articulatory terms, this can be stated as an "open transition" (Catford, 1988): in heterorganic stop-stop clusters, there must be some period of time between the articulatory closures of the first and second stops. If in a given utterance the onset of C2 toward closure (C2 Onset) starts too early with respect to the release of C1 constriction (C1 Release) for whatever reason, there might be no intervening time between the closures of the two consonants. This is prohibited in this language. The present study tested the hypothesis that, in a C1C2 sequence, the stiffness of the C2 closing phase is modulated online during production to ensure open transition. Decreased stiffness results in longer movement durations in achieving the same displacement, all other things being equal (Byrd & Saltzman, 1998). Therefore, by decreasing stiffness, the C2 closing phase will take longer and the obligatory gap between the consonant closures will be maintained.

The present study used 3D electromagnetic articulography data collected from three native speakers of Moroccan Arabic. Stimuli were heterorganic stop-stop clusters in three word positions: word-initial, medial, and final. Clusters of both place orders (Hardcastle & Roach, 1979) were included (anterior-posterior, "tk", and posterior-anterior, "kt"). Stiffness of the C2 closing phase was calculated as the peak velocity (cm/s) of the receiver on the primary oral articulator of the consonant (Lower Lip for /b/, Tongue Tip for /d, t/, and Tongue Back for /g, k/) during the closing phase of the movement divided by the physical displacement (cm) of the receiver during the closing phase (Munhall et al., 1985). Relative C2 Onset was calculated (ms) as the timepoint of C1 Release minus the timepoint of C2 Onset. Larger values indicate earlier Relative C2 Onsets (Figure 1).

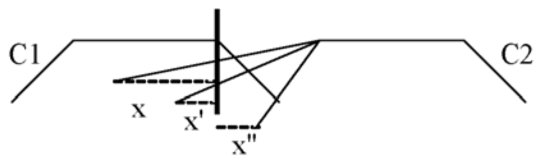


Figure 1: C2 Onset relative to C1 Release (vertical line). Larger positive values (x compared to x') indicate an earlier C2 Onset relative to C1 Release. Negative values indicate C2 Onset after C1 Release (x'').

Pearson correlation coefficients were calculated between Relative C2 Onset and stiffness of C2, within speaker across all word positions. The results strongly supported the hypothesis that stiffness was lower when Relative C2 Onset was larger (i.e., earlier). There was a significant negative correlation for each speaker's data (coefficients: Speaker OB = -0.794, Speaker YZ = -0.789, Speaker CZ = -0.568; for all 3 speakers $p < 0.001$). These correlations were not affected by word position. Scatterplots of each speaker's data are shown in Figure 2. Similar analyses using peak velocity during the C2 closing phase instead of stiffness showed no significant correlations, in further support of the hypothesis that it is stiffness specifically that is modulated.

This result gives insight into one mechanism by which separation of the two closures is effected, and at the same time supports the possibility of obligatory release as a linguistic goal. The question then arises as to whether that goal is articulatory or acoustic. Detailed analysis of these speakers' acoustic and articulatory data is included to address this question.

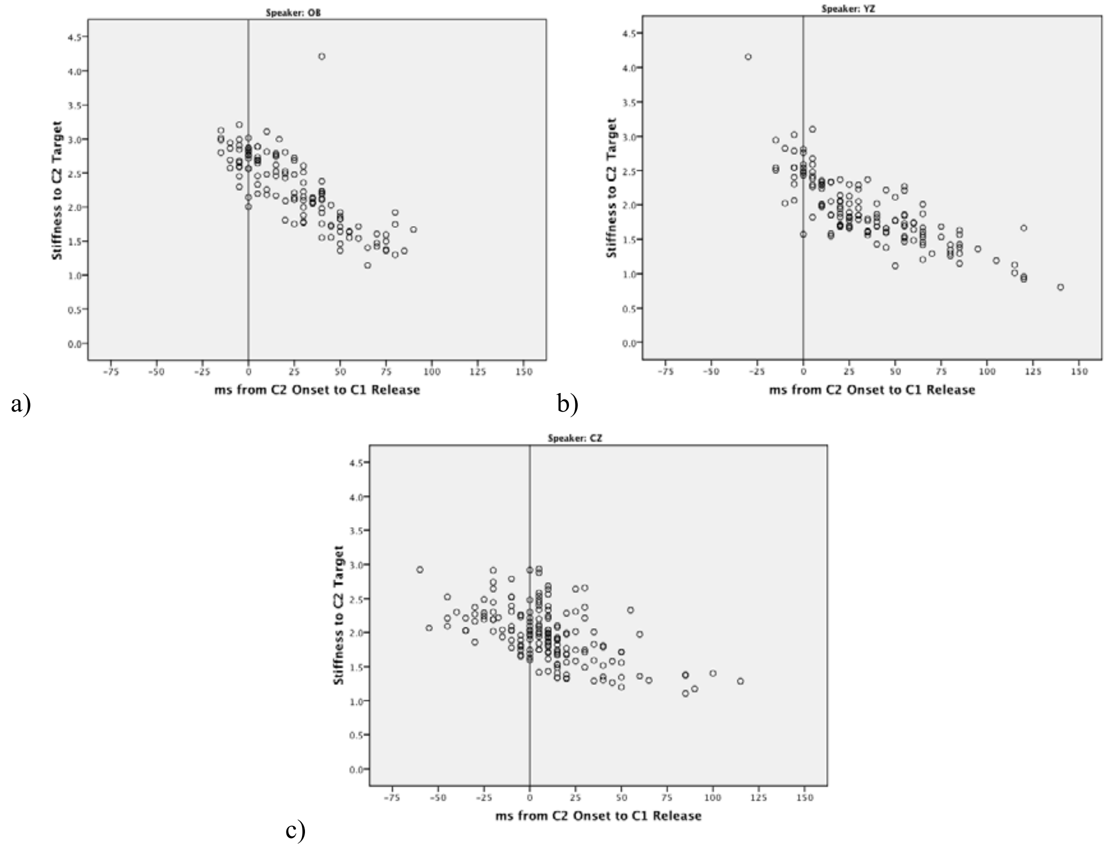


Figure 2: Scatterplots showing the significant negative correlation between Relative C2 Onset with stiffness of C2 closing phase for Speakers OB (a), YZ (b), and YZ (c).

References

Byrd, D., & Saltzman, E. L. (1998). Intragestural dynamics of multiple prosodic boundaries. *Journal of Phonetics*, 26, 173-199.

Catford, J. S. (1988). *A practical guide to phonetics*. Oxford: Clarendon Press.

Dell, F., & Elmedlaoui, M. (2002). *Syllables in Tashlhiyt Berber and in Moroccan Arabic*. Dordrecht: Kluwer.

Hardcastle, W., & Roach, P. (1979). An instrumental investigation of coarticulation in stop consonant sequences. In H. Hollien, & P. A. Hollien (Eds.), *Current Issues in the Phonetic Sciences* (pp. 531–540). Amsterdam: John Benjamins.

Munhall, K. G., Ostry, D. J., & Parush, A. (1985). Characteristics of velocity profiles of speech movements. *Journal of Experimental Psychology: Human Perception and Performance*, 11, 457-474.

Having your cake and eating it: An articulatory perspective on the individual's place in systems of variation and change

James M Scobbie* Jane Stuart-Smith# & Eleanor Lawson⊗

*Speech Science Research Centre, Queen Margaret University; jscobbie@qmu.ac.uk
#Dept of English Language, University of Glasgow; j.stuart-smith@englang.arts.gla.ac.uk
⊗Speech Science Research Centre, Queen Margaret University; elawson@qmu.ac.uk

Scottish English is often cited as a "rhotic" dialect of English. However, researchers have for some time suggested that postvocalic /r/ is in attrition in some varieties of urban Scottish English; notably in the two most populous cities Glasgow (Macafee 1983; Stuart-Smith 2003) and Edinburgh (Romaine, 1978; Speitel and Johnston, 1983; Johnston, 1997). The earliest researchers to note postvocalic /r/ loss were careful to differentiate between Anglo-influenced nonrhoticity, adopted by middle and upper-class speakers (involving straightforward adoption of a nonrhotic pronunciation with associated Anglo vowels), and an apparently local development, confined to lower working-class speech, which involved the weakening of /r/, and it is the latter which we focus on here. Words with weakened /r/ were often transcribed with pharyngealised vowels indicating the retention of a root retraction tongue gesture and hypothesised loss of the tip-raising gesture.

Now it appears that a continuum of local variation has emerged: while some WC speakers have pharyngealised vowels, others apparently no longer produce any form of consonantal /r/ (Stuart-Smith 2003). Moreover, while the non-rhotic and derhoticised variants tended to be used by young working-class Glaswegians and rhotic variants by young middle-class Glaswegians, anecdotal evidence suggests that the derhoticised pronunciations might be spreading up the social scale.

Impressionistic auditory analyses and acoustic analyses of the weakening and loss of /r/ have proved only partly satisfactory, because in an auditory analysis, there is a tendency for listeners (even trained ones) to categorise the /r/ variants. Stuart-Smith (2007) showed that, although there was internal consistency in her transcriber's work, identification of forms as being rhotic or nonrhotic was not consistent between transcribers. This may be expected from a sound as articulatorily complex as /r/ which is undergoing a gradient weakening process, but it would be preferable both to have data which is more objectively gradient, and to investigate the relevance of categorisation for the community itself.

In this paper, we report fieldwork-based and laboratory-based instrumental data on derhoticisation in Scottish English. Acoustic analysis provides a more objective empirical basis for the study of this phenomenon, but on its own, it is both unable to reveal the relative contributions of the various gestures which made up an /r/, and unable to reveal whether there are any covert gestures in which a token which sounds derhotic has in fact a strong inaudible articulation (as suggested by Scobbie et al. 2006). These considerations suggest that a parallel acoustic-articulatory study would be preferable. However, the formal laboratory conditions that usually go hand in hand with instrumental articulatory analysis may cause vernacular informants to "correct" towards rhotic pronunciations due to social pressure towards what is still the Scottish standard, namely rhoticity. A likely candidate for a technique which could capture vernacular articulation is Ultrasound Tongue Imaging (UTI). It is relatively non-invasive and portable, and thus allows data to be collected in informal settings, in a relatively relaxed style.

We therefore have two goals. The first is to test the extent to which UTI interferes with speakers' vernacular style. The second is to explore the articulatory-acoustic relationship in Scottish /r/.

We present data collected under laboratory conditions from a number of Scottish subjects, some of whom have covert pre-pausal tongue blade articulations, i.e. which strong constrictions which do not generate formant movements in low vowels, leading to percepts of derhoticisation (but not phonological neutralisation due to the phonemicisation of allophonic patterns).

The validity of these covert articulations is supported by our methodological fieldwork study. In the latter, fourteen 12-13 year olds were recorded conversing in friendship pairs in their school in Livingston (west of Edinburgh). Informants were recorded first under audio-only conditions and asked to chat alone together in a quiet room for 20 minutes. The following week, five pairs of informants were recorded with audio and UTI and two pairs (experimental controls) were recorded again with audio only. We will report briefly on the UTI fieldwork methodology and the negligible impact of the UTI recording conditions on speech style compared with a repetition of the audio only recording.

Taking both studies together, it appears that weak syllables and utterance-final position are two important conditioning factors for derhoticisation. It is hard to tell with the relatively slow sampling rate of UTI what is happening in the former case, but gestural reduction is compatible with our observations. In the latter, as noted above, we have observed that some speakers have covert anterior articulations: i.e. blade raising which is delayed to a point at which voicing has already ceased, making the constriction barely audible at best, leaving a largely monophthongal vowel.

There are two discussion points. First, the articulatory work which has been carried out previously on American /r/ demonstrates multiple simultaneous constrictions which can vary yet produce similar acoustic outputs (Delattre & Freeman, 1968; Alwan, A & Narayanan, S., 1996; Guenther et al., 1999). Previous UTI work on /r/ focussed on a relatively stable and heterogenous pool of US English speakers (Mielke, Twist, and Archangeli, 2006; Mielke, Baker, and Archangeli 2006), thus supporting these findings that articulatory allophony in /r/ appears to have no acoustic consequences. Variable Scottish /r/ on the other hand appears to be associated with almost the opposite situation: a pattern of acoustic social stratification from clear rhoticity to non-rhoticity, with clear anterior articulations present at both extremes. This leads to our second discussion point, which is the behaviour of the speaker-hearer as an active participant or agent in their speech community. It might be thought that we have presented two mismatching sources of evidence for the "real" nature of the speaker-hearer's cognitive system. Instead, we think that for an individual speaker-hearer, the existence of a final "/r/" as a phonological category (in a given sociolinguistic context) is moot. Our articulatory data supports a more complex and ambiguous conception of "target" within an individuals' output system, reaching across levels, adding to their knowledge of socially-structured variation in the input.

Acknowledgements

We acknowledge the financial support of ESRC Research Grant RES-000-22-2032 and SHEFC SRIF-3.

Selected References

Alwan, A. & Narayanan, S. (1996). Towards articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part II. The rhotics. *Journal of the Acoustical Society of America*, 101, 1078–1989.

Delattre, P. & Freeman, D.C. (1968). A dialect study of American r's by x-ray motion picture. *Linguistics*, 44, 29–68.

Guenther, F. H., Espy-Wilson, C. Y., Boyce, S. E. Matthies, M. L. Zandipour, M., & Perkell, J. S. (1999). Articulatory tradeoffs reduce acoustic variability during American English /r/ production. *Journal of the Acoustical Society of America*, 105, 2854–2865.

Mielke, J., Baker, A., & Archangeli, D. (2006). Covert /r/ allophony in English: variation in a socially uninhibited sound pattern. Paper presented at *Laboratory Phonology 10*, Paris.

Mielke, J., Twist, A., & Archangeli, D. (2006). Are "covert" /r/ allophones really indistinguishable? *NWAV* 35, Ohio State University.

Romaine, S. (1979). Postvocalic /r/ in Scottish English: Sound change in progress? In P. Trudgill. (Ed.), *Sociolinguistic Patterns in British English* (pp. 145–157). London: Edward Arnold.

Scobbie, J. M., Sebregts, K., & Stuart-Smith, J. (2006). From subtle to gross variation: an ultrasound tongue imaging study of Dutch and Scottish English /r/. Paper presented *Laboratory Phonology 10*, Paris.

Speitel, H. H., & Johnston, P. (1983). *A Sociolinguistic Investigation of Edinburgh Speech*. ESRC End of Grant Report.

Stuart-Smith, J. (2003). The phonology of Modern Urban Scots. In G. Corbett et al. (Eds.), *Edinburgh Student Companion to Scots* (pp. 110–37). Edinburgh: Edinburgh University Press.

Stuart-Smith, J. (2007). A sociophonetic investigation of postvocalic /r/ in Glaswegian adolescents. In Jürgen Trouvain & William J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 1449-1452). Saarbrücken, Germany.

The time course of stress and syllable-level encoding: Evidence from a delayed naming task in Spanish

Michael Shelton*, Chip Gerfen# & Nicolás Gutiérrez Palma%

*Department of Spanish and French Studies, Occidental College, USA; mshelton@oxy.edu
#Department of Spanish, Italian and Portuguese, Penn State University, USA; gerfen@psu.edu
%Department of Psychology, University of Jaén, Spain; ngpalma@ujaen.es

The representation of phonotactic knowledge in the lexicon is a long-disputed topic. Traditional approaches to the modeling of syllable weight, for example, posit a categorical representation of a binary distinction between "heavy" and "light" syllables. Evidence from experimental studies, however, has suggested that our intuitions of the possible combinations of sound sequences in our language are much more finely-grained than can be accounted for by abstract categories. These stochastic approaches to the lexicon challenge traditional models by arguing that speakers rely heavily on statistical generalizations across the lexicon when processing phonotactic sequences.

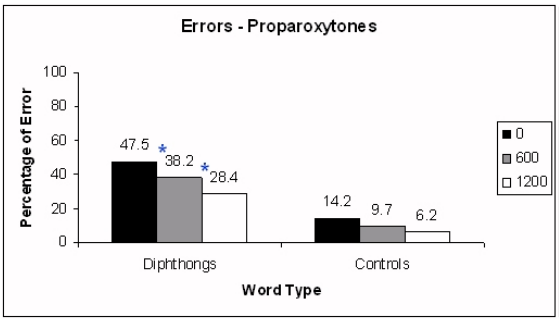
While there is a growing body of literature examining the concept of syllable weight via both theoretical and experimental approaches, the time course of phonotactic processing is an area that has received less attention. Psycholinguistic research on speech production (e.g. Levelt et al., 1999; Dell, 1986) has developed useful tools to test the nature of phonological processing. However, these studies focus most frequently on the cognitive implications of linguistic processing rather than on the insights such methods may lend to the nature of the phonological grammar itself. The goal of the present study is to examine the time course of the processing of stress and syllable-level encoding in Spanish in order to test further the extent to which the interaction of these two levels of phonotactic knowledge may inform the nature of phonological representation within the lexicon.

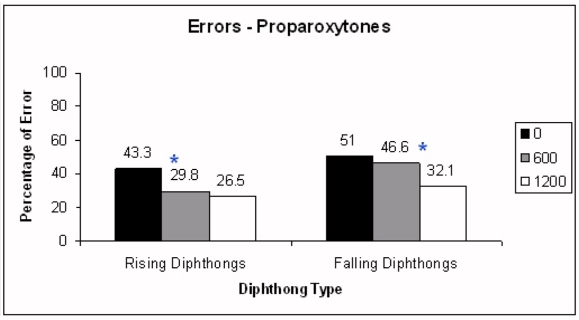
As stress assignment in Spanish is highly constrained by syllable structure, the interaction of the two proves a useful tool for examining phonotactic processing. Briefly, primary stress in Spanish always falls on one of the last three syllables of a word. However, heavy penults restrict stress assignment to one of the final two syllables (Harris, 1983). Though Spanish lacks proparoxytones with either rising (RD) or falling (FD) diphthongs in the penult, these diphthong types do not always pattern together with respect to putative weight equivalence. For example, final RDs regularly fail to attract stress ([fa.mí.lja]), while final FDs systematically pattern as "heavy," attracting stress ([ka.rái]). In our experiment, we employ a delayed nonword naming task to examine the relative strength of the ban in Spanish on RDs versus FDs in penultimate position of an antepenultimately stressed (proparoxytone) word, with the goal of inducing production errors when participants attempt to pronounce illicit word shapes (e.g. *[dóbaina]/*[dóbjana]). In the delayed naming task participants are presented with words on a computer screen and instructed to prepare, but withhold, speech until cued to articulate after a certain delay. This task serves as a useful tool to investigate the processing of phonological information over time in that varying delays offer controlled amounts of preparation for phonological processing.

Our prediction is this: if the representation of phonological structure is categorical, the time required to process RDs and FDs should be relatively equivalent, given the "heavy" nature of both RDs and FDs within traditional phonological frameworks. However, a difference in error rates between diphthong types across delay conditions would indicate that the manipulation of delay time reveals more subtle patterns in phonological processing, suggesting that speakers are sensitive to very subtle patterns across the lexicon.
55 monolingual Spanish undergraduate students at the University of Jaén, Spain named 112 critical three-syllable nonce forms, broken down into 1) nonce proparoxytones with a falling diphthong (FD) in the penult (e.g. *dótaiga*), 2) nonce proparoxytones with a rising diphthong (RD) in the penult (e.g. *dóbiana*), and 3) nonce CV.CV.CV controls (e.g. *dóveno*). Additionally, each subject saw 168 filler items consisting of 28 additional nonce forms and 140 real words divided into categories of oxytone, paroxytone and proparoxytone stress of high and low frequency. Stimuli were presented in E-Prime, and subjects named each item as quickly and accurately as possible after cued to speak by a tone which sounded 0ms, 600ms, or

1200ms after the presentation of the word. Responses were recorded digitally for subsequent coding of error rates. Critical proparoxytone nonce forms also had segmentally controlled paroxytone counterparts. Stimuli were randomized and counterbalanced across subjects, i.e. if a subject saw *dóveno*, s/he did not also see *doveno*.

The prohibition on antepenultimate stress in words containing a diphthong in the penultimate syllable was clearly present in our error rate data. Diphthong stimuli evoked significantly more errors than controls across all delays ($p<0.01$) and showed significant improvement with each delay ($p<0.01$). Most interestingly, however, is the finding of inconsistent improvement of diphthong type across delays. RDs show significant improvement in production with 600ms of preparation time (600ms=$p<0.02$, 1200ms=$p>0.99$), whereas FDs require twice as long to show significant improvement (600ms=$p>0.98$, 1200ms=$p<0.01$).





The comparison graph on the left clearly illustrates the phonological prohibition on antepenultimate stress with a diphthong in the penultimate syllable, as the arguably "heavy" penults induce more errors than the licit "light" controls. It is also clear that as more time is given for preparation speakers are able to improve their production of illicit forms. This comparison appears to confirm the traditional division between "heavy" and "light" syllables. However, the second graph, comparing RDs and FDs across delay conditions, reveals a distinction not predicted under traditional phonological models in which both diphthong types are taken to be heavy, thus blocking the possibility of antepenultimate stress.

In these data, rising diphthongs exhibit no additional advantage in the 1200ms condition over the 600ms condition. By contrast, error rates in FDs are significantly lower in the 1200ms condition than in the 600ms condition. We conclude from these data that the phonological processing load of nonword proparoxytones is greater for the FD than the RD forms.More generally, our data provide evidence from a speech production task that speakers are highly attuned to subtle phonotactic patterns in the lexicon when planning and producing Spanish—patterns that represent finely-grained differences not predicted by the categorical division between "heavy" and "light" syllables in traditional phonological models. This is not to say that these coarse generalizations are not also represented in the grammar, but rather that speakers are able to generalize across the lexicon at multiple levels of abstraction. Thus, in one sense, our diphthongs/controls comparison reflects the phonotactic ban on antepenultimately stressed words with a diphthong in the penultimate syllable. At the same time, speakers exhibit a striking sensitivity to subtle differences in the strength of the proscription, differences arguably driven by the distinct patterning of rising and falling diphthongs with respect to stress in other phonological contexts across the lexicon. The manipulation in our delayed naming task provides a means of teasing out this granularity in an on-line behavioral measure.

References

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review,* 93, 283-321.

Harris, J. W. (1983). *Syllable structure and stress in Spanish: A nonlinear analysis.* Cambridge, MA: MIT Press.

Levelt, W. J. M., Roelofs, A., & Meyers, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22, 1-38.

The Acoustics of Voiceless Nasal Vowels

Ryan K. Shosted

Department of Linguistics, University of Illinois, Urbana-Champaign;
rshosted@uiuc.edu

Introduction

Can nasal vowels, like oral vowels, devoice? While phonemic voiceless nasal consonants are found in a variety of languages, phonemic voiceless nasal vowels appear unattested (Ladefoged & Maddieson, 1996; Crothers et al., 1979). Typological evidence also suggests that phonological processes of vowel devoicing eschew nasal vowel targets. Of the 55 languages with voiceless vowels or vowel devoicing processes catalogued by Gordon (1998), only four of these also have phonemically nasal vowels. These are Bagirmi, Montreal French, Mbay, Mixtec, and Brazilian Portuguese. While it is clear that nasal vowel devoicing is impossible in French and Portuguese, a review of the other languages reveals no positive evidence that it is possible (or impossible) for nasal vowels to be realized without vocal fold vibration.[1]

This paper asks whether acoustic factors may present barriers to the perception of voiceless nasal vowels. A series of perceptual experiments using whispered nasal and oral vowels of Brazilian Portuguese will test whether "devoiced" nasal vowels are more difficult to identify than "devoiced" oral vowels. The acoustic characteristics of whispered nasal vowels will also be examined to see if they provide any clues as to the typological rarity of nasal vowel devoicing.

Perception Experiment

A female speaker of Brazilian Portuguese was recorded pronouncing oral/nasal minimal pairs in whispered and modal voice. Since nasal vowel devoicing is impossible via phonological rule in Brazilian Portuguese, whispered voice was used as a surrogate for voicelessness. During whisper, the vocal folds assume a slightly different configuration than during voicelessness. However, both phonation types share a lack of vibration at the glottis, the most relevant feature for the present study. All tokens had either word-final stress or were monosyllabic and terminated with either a nasal or oral vowel.

In Experiment 1, the final vowel was excised and a Hamming taper was applied. Two stimulus classes resulted from this process: natural whispered nasal and natural whispered oral tokens. Using Burg's LPC method, 16-coefficient acoustic filters were created for all the vowels. These filters were then applied to a white noise source to produce four classes of synthesized tokens: whispered nasal, whispered oral, modal nasal, and modal oral. The final two classes were used as controls. In Experiment 2, whole word tokens were excised and amplitude-normalized. Three native listeners of Brazilian Portuguese participated.

In the first experiment, male and female subjects listened to 276 natural and synthesized vowel tokens along with fillers. The tokens were presented in randomized order using E-Prime. Each vowel extract was played for the subject once, and the subject was asked to choose whether s/he heard /a e i o/ or /u/. In the second experiment, subjects heard whispered versions of 20 oral/nasal minimal pairs, accompanied by fillers (832 stimuli in total). Subjects heard each stimulus and chose whether it ended in a nasal or oral vowel.

Perceptual sensitivity and response bias were calculated (Macmillan & Creelman, 2005). Experiment 1 suggests that perceptual sensitivity to vowel category is degraded by nasality in the absence of a voicing source and that the perception of mid-vowels /e o/ suffers more than the perception of the corner vowels /a i u/. Experiment 2 suggests that the nasal/oral distinction is perceptually robust in whispered vowels. Based on the Acoustic Theory of Speech Production (Fant, 1960) and changes to F1 predicted by nasal coupling, it is tempting to wonder whether the distinction between oral and nasal whispered vowels is most salient for vowels with maximal or minimal F1, /a i u/, and less salient for those vowels with intermediate F1 values, /e

[1] Nasal vowel devoicing is impossible in Brazilian Portuguese only because nasal high vowels do not occur in unaccented final position (note, however, that non-high nasal vowels can occur in unaccented final position, e.g. *órfã* 'orphan.FEM'). In Montreal French, nasal vowel devoicing is impossible because only high vowels devoice and there are no high nasal vowels. Thus, a proper test case remains elusive.

o/, as observed in Experiment 1. This result, however, did not obtain in Experiment 2.

Acoustic Measures

Whispered oral, whispered nasal, modal oral, and modal nasal vowels were compared to one another with regard to their duration. While voiced nasal vowels are typically longer than voiced oral vowels (Whalen & Beddor, 1989), this distinction appears to be lost when the pairs are whispered.

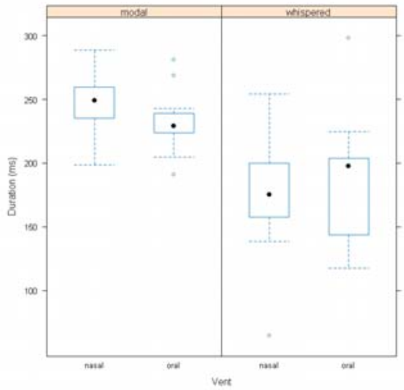


Figure 1: Boxplots show the duration of whispered versus modal, nasal and oral vowels in Brazilian Portuguese.

Conclusion

Despite the fact that voiceless nasal vowels are not difficult to produce,[1] they are uncommon—if not entirely unknown—in phonological systems. This study shows that the quality of whispered vowels is more difficult to distinguish if the vowel is nasal than if it is oral. It also shows that duration contrasts, important for cueing the difference between oral and nasal modal vowels, may be lost when the vowels are whispered. To the extent that whispered vowels can provide a window to the acoustics and perception of voiceless vowels, these results suggest an acoustic-perceptual explanation for the absence of voiceless nasal vowels.

References

Crothers, J., Lorentz, J., Sherman, D., & Vihman M. (1979). *Handbook of Phonological Data from a Sample of the World's Languages*. Palo Alto, CA: Department of Linguistics, Stanford University.

Fant, G. (1960). *Acoustic Theory of Speech Production*. The Hague: Mouton.

Gordon, M. (1998). The phonetics and phonology of non-modal vowels: A cross-linguistic perspective. In *Proceedings of the Berkeley Linguistics Society*, 24, 93–105.

Ladefoged, P., & Maddieson, I. (1996). *The Sounds of the World's Languages*. Oxford: Blackwell.

Macmillan, N. A., & Creelman, C.D. (2005). *Detection Theory: A User's Guide* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Ohala, J. J. (1983). The origin of sound patterns in vocal tract constraints. In P. F. MacNeilage (Ed.) *The Production of Speech* (pp. 189-216). New York: Springer Verlag.

Whalen, D. H. & Beddor, P.S. (1989). Connections between nasality and vowel duration and height: Elucidation of the Eastern Algonquian intrusive nasal. *Language*, 65, 457–486.

[1] Ohala (1983: 202-206) has argued that vowel devoicing is an aerodynamic outcome relating to tongue constriction and concomitant increase in back pressure. For this reason, he argues, high vowels are common targets. Because pressure is continuously vented during the production of nasal vowels, one might well agree with him that there is little likelihood of a nasal vowel undergoing epiphenomenal devoicing. However, while high-vowel devoicing is clearly preferred, there are still languages (like Tongan) that devoice low vowels despite a lack of supraglottal pressure increase (Gordon 1998: 98, ff. 106). Thus, one cannot rule out voiceless nasal vowels on aerodynamic grounds alone.

Phrasing not duration differs in French questions and statements

Caroline L. Smith

Department of Linguistics, University of New Mexico; caroline@unm.edu

In the majority of languages, the most salient prosodic difference between statements and questions is their distinct intonational patterns. Many languages allow statements and questions that are syntactically identical and are distinguished solely by intonation. However, there have been suggestions that the two sentence types may differ in their durational patterns. Van Heuven and van Zanten (2005) propose as a possible universal that questions are produced with a faster speech rate than statements. In their study of Manado Malay, Orkney English, and Dutch, they found a faster rate for questions in all three languages. Faster speech rate in questions has also been reported for Canadian French, but the increase in rate was accompanied by more final lengthening in questions than in statements (Ryalls, Le Dorze, Lever, Ouellet, & Larfeuil, 1994; Smith, 2002). The present study investigates timing in French statements and questions in order to determine if, and in what way, their durations pattern differently over the course of an entire sentence.

Six native speakers of metropolitan French were recorded reading sets of statements and questions for an acoustic study of durational patterning. In order to maximize comparability, ten sets of sentences were created in which the sentences in each set were as similar as possible except for varying sentence type. Each set included a statement and a lexically identical, syntactically-unmarked “Echo” question. Sets also included three other sentences that included many of the same lexical items as the first two: a question with subject-verb inversion (“Inverted question”), a question introduced by a question word such as *qui*, “who” (“Partial question”), and another statement in which the word that was sentence-final in the other lists was sentence-medial (“Control statements”). The speakers read the sentences six times from six different lists. The lexically identical pairs of statements and Echo questions appeared together as the first part of the list. These were grouped in pairs so that the statements provided a context for the Echo questions, which the speakers were instructed to read as if they could not believe the preceding statement. As the second part of the list, all of the Inverted questions, Partial questions and Control statements were randomized together, in a different order in each list. The Partial questions were not analyzed in this study.

Speakers were recorded individually in Lyon, France, and the recordings digitized at 10kHz on a Kay Elemetrics CSL system. All measurements reported here were obtained using Praat (Boersma & Weeenink, 2007). Results are reported here for four speakers, two male and two female, and for three of the sets of sentences. These three sets were chosen out of the ten recorded because they had the greatest lexical similarity among the sentences within the sets, and within each set, the same number of syllables across the different sentences. All of the sentences were divided into syllables, with the segmentation consistent across all repetitions by all speakers. This meant that if one or more speakers made a break between a consonant-final word and a vowel-initial word, the consonant was always syllabified with the preceding syllable, as for example, the [m] in *ce co<u>stume e</u>st-il* … [stym . (ʔ) e .], ‘Is this costume …’. (If followed immediately by a vowel, the consonant would otherwise normally syllabify as the onset of the following syllable.)

Speech rate was calculated in syllables per second over entire sentences, excluding silent intervals. Only the Inverted questions and Control statements were used for calculations of rate. These occurred in random order, with sentences from different sets interspersed with each other. Rate could not be compared in the statements and Echo questions, because these were always presented with the statement immediately preceding the Echo question, and the repetition of identical material meant that speakers invariably increased rate in the Echo question. Contrary to the findings of Ryalls et al. (1994), all four speakers in this study showed a consistent trend for faster speech rate overall in statements (5.91 sylls/sec) than in questions (5.80 sylls/sec), although the effect did not reach significance ($t(11)=1.57$).

The statements and questions were divided into phrasal groups in order to identify which sentence-medial word(s) occurred at the end of a phrase. The phrasal units used in this study correspond approximately to the tonally-defined Accentual Phrase of Jun and Fougeron (2002). The final words, and particularly the final full-vowel syllables, in these phrasal groups are subject to lengthening, but less so than the words at the end of larger phrases. In order to control for differences in overall rate and intrinsic syllable durations, the variable that was examined was the ratio of the duration of the (sentence-medial) syllable at the end of the first Accentual Phrase in the sentence to the duration of the sentence-final syllable. The larger that this ratio is, the more the sentence-medial syllable has been lengthened. Analyzing all four speakers together, the ratio was significantly larger in the statements than in Echo questions ($t(11)=4.45$, $p<0.001$) and in the statements compared to the Inverted questions ($t(11)=3.04$, $p=0.01$). Thus the statements have greater lengthening at the end of the sentence-medial phrase than do the questions.

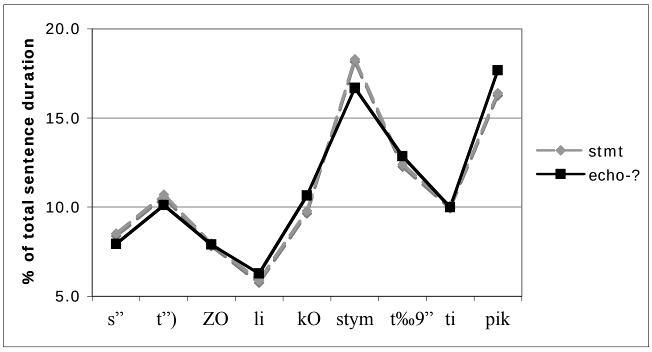


Figure 1: Percentage that each syllable contributes to total sentence duration in lexically identical statement and echo question *C'est un joli costume très typique*. 'It's a pretty, very typical costume.'

This pattern is very apparent in Figure 1, which shows the average of six repetitions of two sentences by one speaker, but illustrates a pattern found consistently. It shows the percentage of total sentence duration taken up by each syllable. The percentages are almost identical except for the phrase-final syllables, which are [stym] at the end of the first phrase and [pik] at the end of the sentence.

Although the statements and Echo-questions were syntactically identical, the speakers phrased them differently (as expected, see Wunderli 1984). All of the words and syllables identified as phrase-final occurred at the end of a phrasal group in every production of the statements. Although the questions and statements were matched in length, these words were judged by the experimenter to be phrase-final in only 37 of 72 productions of Echo questions, and 42 of 72 Inverted questions. Thus a likely explanation for the greater lengthening found sentence-medially in statements is that these words were consistently phrase-final, but many of the words measured in questions were not at the end of any phrasal unit. Thus the durational difference most likely is explained by the fact that speakers frequently produced both types of questions as a single phrasal unit, whereas all of the statements were produced as two or sometimes three phrases. The different numbers of phrases do not explain the greater lengthening found sentence-finally in questions, unless the number of syllables in a phrase contributes to determining the amount of lengthening at the end of the phrase. These results suggest that overall rate differs little between statements and questions; the difference lies in the distribution of lengthening within the sentence. Questions may appear to have a faster rate, but this actually reflects a difference in phrasal structure, not an intrinsic difference in speaking rate.

References

Boersma, P., & Weenink, D. (2007). Praat: doing phonetics by computer (Version 4.6.39). http://www.praat.org.

Jun, S.-A., & Fougeron, C. (2002). Realizations of accentual phrase in French intonation. *Probus*, 14, 147-172.

Ryalls, J., Le Dorze, G., Lever, N., Ouellet, L., & Larfeuil, C. (1994). The effects of age and sex on speech intonation and duration for matched statements and questions in French. *Journal of the Acoustical Society of America*, 95, 2274-2276.

Smith, C. (2002). Prosodic finality and sentence type in French. *Language and Speech*, 45, 141-178.

van Heuven, V., & van Zanten, E. (2005). Speech rate as a secondary prosodic characteristic of polarity questions in three languages. *Speech Communication*, 47, 87-99.

Wunderli, P. (1984). L'intonation des questions sans marque segmentale. *Travaux de Linguistique et de Littérature*, XXII, 203-250.

Bringing Semantics to Sociophonetics: Social Variables and Secondary Entailments

E. Allyn Smith,* Kathleen Currie Hall,* & Benjamin Munson#

*Department of Linguistics, The Ohio State University, USA; esmith, kchall @ling.osu.edu
#Department of Speech-Language-Hearing Sciences, University of Minnesota, USA; Munso005@umn.edu

One of the questions being addressed in the 'social information in the lexicon' sub-theme of LabPhon 2008 is whether phonetic information in the lexicon is accompanied by social information. The lexicon is generally thought of as the nexus for stored information about the form and meaning of linguistic elements. Traditionally, studies examining social information and the lexicon have focused on whether individuals can infer social-group membership from distinctive patterns of phonetic variation, and whether social expectations affect phonetic labeling. The results of numerous studies show that both are indeed the case (Purnell et al., 1999; Hay et al., 2006; *inter alia*). However, we argue that a different line of research is needed to fully address whether social information is represented in the lexicon in a manner analogous to how "regular" linguistic meaning is represented.

Specifically, we argue that if social information is in fact a type of linguistic meaning, it should be treated in the same way as other kinds of meaning. In this paper, we describe two primary distinctions that formal semanticists use to categorize meaning (first, conventionality, and second, whether it responds to the question under discussion) and illustrate how at least one social variable that is instantiated in the phonology can be described in this framework as one of these traditional semantic categories. The social variable we use to illustrate our approach is the low-back/central production of /æ/ as [æ̠] in Minnesota that is associated with judgments of gay-soundingness in speech (Munson et al., 2006), and we compare this variable to a *secondary entailment* (such as a Conventional Implicature, cf. Potts, 2005). Here, we describe the semantic tools used to determine secondary entailments and show how they can be extended to social variables.

The first division that semanticists generally make is between *conventional* and *conversational* types of meaning. Conventional types of meaning are robust to different conversational contexts, and must be part of the stored grammatical knowledge of language users. For example, compare *know* with *discover*. Both are said to carry "factive presuppositions," meaning that the truth of their sentential complement is required (as in (1) and (2)). However, the factive presupposition of *know* is conventional, while *discover*'s is conversational (Abusch, 2002). We verify this difference by showing that in a certain context (such as the antecedent of a conditional in (3) and (4)), *discover*'s presupposition does not go through, while *know*'s does.

(1) Joe discovered that Nia is in New Zealand. (Nia must be in NZ)
(2) Joe knows that Nia is in New Zealand. (Nia must be in NZ)
(3) If Joe discovers that Nia is in New Zealand, I'll be mad. (Nia is not necessarily in NZ)
(4) If Joe knows that Nia is in New Zealand, I'll be mad. (Nia must be in NZ)

The second division that semanticists make is between different types of conventional meaning: there are *primary* ("at-issue") entailments and *secondary* ("not-at-issue") entailments (Potts, 2005). These are divided based on whether they respond to the Question Under Discussion (QUD) (Roberts, 1996). If they do at least partially respond to the QUD, they are at-issue, and if not, they are not-at-issue. For example, in (5), the information provided by *William Taft had a moustache* answers the QUD, while the information provided in the appositive (*previously the Governor-General of the Philippines*) does not. Thus, the appositive is considered secondary, while the rest is primary. Both are entailments because the sentence cannot be judged to be true unless both of these parts are also true.

(5) QUD: Which United States Presidents had moustaches?
A: William Taft, previously the Governor-General of the Philippines, had a moustache.

Consider now how these semantic distinctions can be applied to a social variable. In Minnesota, a pronunciation of /æ/ as [æ̠] is generally perceived as a marker of gayness. For example, (6), with *mad* pronounced as [mæ̠d], is traditionally associated with gay speech. If social meanings are secondary entailments, the fact that the dog is mad is primary, and the fact that the speaker is gay is secondary.

(6) A: That dog is really mad. (7) B: No, it just looks mad.
(8) B: # No, you're not gay. (9) B: Hey wait a minute! I didn't know you were gay...

One method of testing for secondary entailment is the *Hey wait a minute!* (HWAM) test (Shannon 1976). Once something is asserted, its primary entailment can be called into question by beginning a sentence with *no*, while its secondary entailment cannot; however, it can be called into question using HWAM. Native speaker intuitions show that (7) and (9) but not (8) are appropriate answers to (6), indicating that the HWAM test, when applied to [æ̠], supports the hypothesis that [æ̠] is a secondary entailment. In this paper, we show how the social meaning associated with [æ̠] patterns in comparison to other secondary entailments with respect to a variety of semantic tests, including HWAM, speaker-orientation, strengthening, and invariance under presupposition holes and plugs. To the extent that the phonetic variable [æ̠] in Minnesotan English patterns like other secondary entailments, the association between [æ̠] and its social meaning is conventional, and thus part of the grammar of Minnesotan English speakers.

To answer the question of whether this association is stored in the lexicon, however, we must have a precisely developed definition of the lexicon. In an exemplar-based model of grammar (e.g. Pierrehumbert 2001), semantic and social information may either be stored in the repository of exemplars itself (e.g, an exemplar could contain multiple pieces of information, such as {[kæ̠t], N, **cat′ • gay′(speaker)**}) and thus be "in the lexicon," or such non-phonetic information could be stored in separate modules, and linked to phonetic representations in the lexicon (e.g., an exemplar could contain {[kæ̠t]} and then be linked to a semantic module containing the meaning **cat′ • gay′(speaker)**). Thus the decision about whether phonetic information *in the lexicon* is accompanied by social information will be dependent on one's choice of defining the lexicon. Similarly, in a traditional lexicon-plus-generative-rules approach to grammar, it is possible to store social meaning with each individual lexical item (e.g., {[kæ̠t], N, **cat′ • gay′(speaker)**}) or to encode the social meaning as part of a multi-modal rule (e.g., /æ/ → [æ̠] iff **speaker → gay′(speaker)**).

What is important here, however, is that the tests developed by semanticists to determine and categorize "regular" semantic meaning can be applied to social meaning, and any model of the grammar must allow phonetic information to be accompanied by social information in the grammar itself, regardless of whether it appears in the lexicon *per se*.

References

Abusch, D. (2002). Lexical alternatives as a source of pragmatic presuppositions. In B. Jackson (Ed.), *Proceedings of SALT XII*. Ithaca, NY: CLC Publications.

Hay, J., Warren, P., & Drager, K. (2006). Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics,* 34(4), 458-484.

Munson, B., McDonald, E. C., DeBoe, N. L., & White, A. R. (2006). Acoustic and perceptual bases of judgments of women and men's sexual orientation from read speech. *Journal of Phonetics,* 34, 202-240.

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition, and contrast. In J. L. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 137-157). Philadelphia: John Benjamins.

Potts, C. (2005). *The logic of conventional implicatures*. New York: Oxford University Press.

Purnell, T., Idsardi, W. J., & Baugh, J. (1999). Perceptual and phonetic experiments on American English dialect identification. *Journal of Language and Social Psychology,* 18(1), 10-30.

Roberts, C. (1996). Information structure in discourse: Towards an integrated formal theory of pragmatics. In J. Yoon & A. Kathol (Eds.), *OSU working papers in linguistics* (pp. 91-136). Columbus, OH.

Shannon, B. (1976). On the two kinds of presuppositions in natural language. *Foundations of Language,* 14, 247-249.

Voicing control and nasalization

Maria-Josep Solé*, Ronald Sprouse & John J. Ohala#

*Department of English, Universitat Autònoma de Barcelona, Spain; mariajosep.sole@uab.es
#Department of Linguistics, University of California, Berkeley; ronald@berkeley.edu; ohala@berkeley.edu

Introduction

The dependency between nasality and voicing has been accounted for by 'redundancy rules' or OT constraints (e.g., NAS/VOI or *NC8), amongst others. This paper presents an account of the physical factors responsible for the dependency between nasality and voicing. In particular, it provides an aerodynamic explanation for cross-linguistic patterns that cannot be accounted for by acoustic-auditory factors (e.g., the lesser tolerance of voiceless stops vis-à-vis voiced stops to nasalization; Ohala and Ohala, 1991), the *NC8 constraint (Pater, 1999), or post-nasal voicing (Hayes and Stivers, 1996). For example, languages with distinctive voiceless stops, [p t k], and prenasalized voiced stops, [^{m}b, ^{n}d, ^{N}g], (illustrated in 1) but no simple voiced stops (suggesting that such prenasalized stops form the voiced stop series); phonetic pre- and postnasalization of voiced (but not voiceless) stops in the absence of contextual nasals (illustrated in 2); the emergence of non-etymological nasals adjacent to voiced but not voiceless stops (see 3) cannot be accounted for by the lesser tolerance of voiceless stops to nasalization or by postnasal voicing simply because these cases do not contain a nasal etymologically or occur in a nasal context. Similarly, the maintenance of the voicing stop contrast exclusively in a nasal context (see 4) cannot be explained by the principles above.

(1) Voiceless vs prenasalized stops, e.g., Waris [^{m}b] banda *'snake'* [p] panda *'pitpit type'*
[^{ŋ}g] go *'go'* [k] kao *'tree sp.'*

(2) Prenasalization, e.g., Bola [b] ~ [^{m}b] bahele *'crocodile'* [d] ~ [^{n}d] dagi *'dig'* [g] ~ [^{ŋ}g] ge '*3rd pers. fem*'
Post-nasalization, e.g.,
Lancashire [ˈspɪt ə ˈgɒbm] *spit a gob* [ˌu:z ə ˈwɛdn] *she's wed* [ˈkɔ: ə ðɪ ˈlɛgŋ] *calf of thy leg*

(3) Child phonology: English [bent] for *bed* [bed]; Greek [^{ŋ}gol] for [gol] *'goal'*
Second language acquisition: Spanish [^{m}baɲo] for *baño* [baɲo] 'bathroom'

(4) Stop voicing contrast preserved only postnasally, e.g.,
Majorcan Catalan /b/ dobl [pl] *'I double'* sembl [bl] *'I think'*
/p/ acopl [pl] *'I fit together'* umpl [pl] *'I fill'*
Basaa pên *'color'* li-pém *'honor'* m-pen *'prong of a fork'* lép *'water'*
– – m-bên *'handle'* –

These patterns suggest that nasals emerge in oral contexts to facilitate voicing in stops. Indeed nasal leakage during the stop closure lowers oral pressure and thus increases the transgottal flow, which favors voicing. Thus nasal leakage is seen as a maneuver to facilitate voicing in stops and, ultimately, to preserve the voicing contrast. This explanation is supported by data from a variety of languages indicating that nasals occur more often in contexts where vocal fold vibration is more difficult to maintain/initiate, for example, in velars vis-à-vis more anterior stops, and in initial vis-à-vis medial position.

Experimental data and modeling

In order to test the hypothesis that nasal leakage may be used to preserve voicing in the stop an experiment was conducted to evaluate the effect of pressure variations with a pseudo-nasal valve on allophonically devoiced stops in American English. Oral pressure was randomly bled during the production of devoiced stops in initial position (e.g. **b**oy), final position (e.g., la**b**), and in stop clusters (e.g., 'la**b** **b**oy'). The values of peak oral pressure and Voice Onset Time (or voicing into the stop closure) for each of the tokens were correlated for each position (initial, final, cluster). The results for utterance initial (e.g., **b**oy, **g**ame) and utterance final (e.g., la**b**, ba**g**) stops (see Fig. 1) show that when oral pressure is reduced by nasal venting, voicing is anticipated in initial stops (note the longer negative VOT values, left panel), and prolonged in final stops (longer voicing into the stop closure, right panel). Nasal venting also prolongs voicing in voiced stop clusters. The results indicate that nasal leakage increases the transglottal pressure drop and facilitates voicing during the stop. The results also indicate that speakers had an adducted (closed) glottis during the production of devoiced stops and that the stops were passively devoiced. When oral pressure was reduced (and thus t

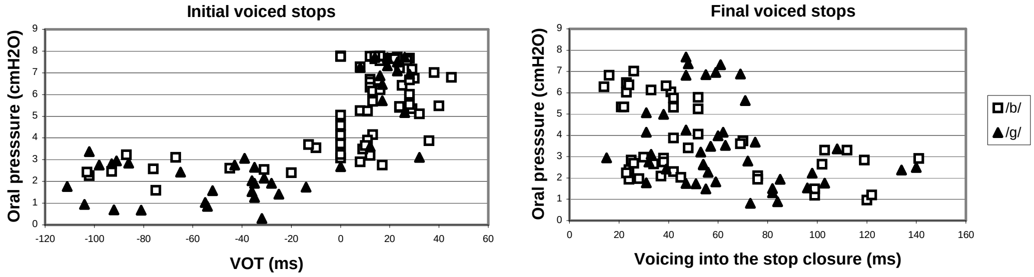


Figure 1: V ... ariation in oral pressure due to nasal venting.

Variations in nasal leakage during the stop closure and their effects on voicing were simulated with Sprouse's (2008) aerodynamic model and are illustrated in Fig. 2. The left panel shows that in utterance-initial stops $P_{subglottal}$ and P_{oral} rise during the closure interval, and the pressure drop is below the critical level for voice-initiation ($4cmH_2O$) before the release of the consonant. In utterance-final stops, voicing extinguishes when the pressure difference drops below $2cmH_2O$, approximately 20ms after the stop closure (without wall compliance). The right panel shows velopharyngeal opening during most of the duration of the initial and final stops and a sufficiently large pressure differential for voicing throughout the stop closures. (For the initial stop the velum starts to close shortly before the stop release so as to produce an oral burst).

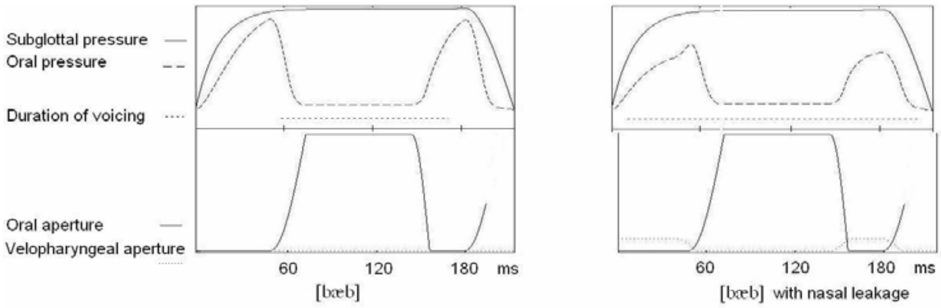


Figure 2: Aerodynamic modeling of the duration of voicing in initial and final stops without and with nasal leakage.

Conclusions

The results of the experiment indicate that reducing the oral pressure increases the rate of flow through the glottis (due to the larger ΔP) which helps maintain/initiate voicing. The results provide support to the explanation that nasal leakage may be exploited by speakers to promote voicing during stops and, ultimately, to maintain the voicing contrast. Such nasal leakage may be reinterpreted by listeners as intended nasals and thus get encoded in the phonology (Ohala 1983). Finally, it is argued (i) the results suggest caution to posit innate constraints to account for the interaction between nasality and voicing when physical phonetic factors have not been discarded, and (ii) that formal phonological notations which represent the nasal valve and the larynx at different nodes fail to capture the interaction between nasality and voicing.

References

Hayes, B., Stivers, T. (1996). A phonetic account of postnasal voicing. Ms., Dept. of Linguistics, UCLA.

Ohala, M., & Ohala, J. (1991). Nasal epenthesis in Hindi. *Phonetica,* 48, 207-220.

Pater, J. 1999. Austronesian Nasal Substitution and other NC effects. In R. Kager, H. van der Hulst & W. Zonneveld (Eds.), *The Prosody-Morphology Interface*(pp. 310-343). Cambridge: Cambridge University Press.

Sprouse, R. (2008). A discrete time aerodynamic model of the vocal tract. *Proceedings of the 8th International Seminar on Speech Production*. Strasbourg.

Vowel length effects on the perception of geminate and singleton stop boundaries by Korean learners of Japanese

Mee Sonu

Graduate School of Japanese Applied Linguistics, Waseda University, Japan;
sonumee@toki.waseda.jp

Introduction

Generally, closure duration is a primary cue used to distinguish plosive geminate and singleton stops (Fukui, 1978). However, closure duration is not an absolute value. It is relative. That is, closure duration is controlled by the neighboring phonemes. However, it is not yet clear what factors affect the closure duration at the perceptual boundary between geminate and singleton stops by either Japanese native speakers or by Korean learner of Japanese.

Accordingly, this paper investigates the relationship between the preceding vowel and the closure duration at the perceptual boundary by Japanese native speakers and Korean learners of Japanese.

Concretely, this paper's research questions are as follows. First, what is the relationship between the preceding vowel length and the perceptual boundary of the geminate and singleton stops? Second, what are the differences, if any, between the perceptual boundary of Korean learners of Japanese and that of Japanese native speakers?

Method

Participants

Eighteen native speakers of Japanese from Tokyo and forty-five Korean Japanese learners participated in the study. They exhibited no symptoms of hearing disability and were compensated for their participation.

Stimuli

The materials were three pairs of 2 mora and 3 mora words which contrasted singleton and geminate stops (/aka/-/akka/, /saka/-/sakka/, /raka/-/rakka). Each item was produced both in isolation and embedded in a carrier sentence: for example, *watasiwa ___to iimasita* (I said ____). A female Japanese speaker spoke the items with an LH-type pitch accent at a normal speaking rate. The materials were digitally recorded.

A stimulus set was created from the materials by acoustically modifying the duration of the previous vowel and the closure of the unvoiced plosive between the first and second mora of each word. Two preceding vowel lengths (PV, long and short) were used and the length of the stop was modified to provide samples varying by 20ms in length by either inserting silence or removing part of the closure.

Procedure

Each subject heard each stimulus a single time and was asked to judge whether he or she heard had heard a 3-mora or 2-mora word. Feedback was not provided during the judgment task. To analyze the results, the perceptual boundary point (B.P.) was defined as the point which measured a 0.5 response rate on the logistic curve for each item between geminate and singleton stops.

Results

The results for the first research question are as follow. The perceptual boundary point is longer when the preceding vowel length was shorter (see Table 1). A repeated measured ANOVA revealed main effects for the length of the preceding vowel [$F(1, 61)=416.0$, $p<0.001$]. This result indicates the perceptual boundary of the geminate and singleton stop significantly depends on the length of the preceding vowel.

The results for the second question are as follows. The B.P was different between the Japanese native speakers and Korean learners of Japanese on all items. The ANOVA also revealed main effects of the preceding vowel between groups [$F(1, 61)=6463.9$, $p<0.001$]. Moreover, the analysis determined that some Korean learners of Japanese were not able to distinguish between geminate and singleton stops.

Table 1: Perceptual boundary point results

		Korean Speakers (N=45)		Japanese Speakers (N=18)	
items	Preceding Vowel (ms)	B.P		B.P	
		mean	SD	mean	SD
/aka/-/akka/	115	112.5	11.5	94.0	14.0
	91	123.7	11.0	109.5	12.7
/saka/-/sakka/	94	92.8	11.8	82.3	12.2
	73	103.9	12.4	95.0	16.0
/raka/-/rakka/	116	93.2	13.3	77.8	13.4
	99	95.6	16.7	88.0	14.5

Discussion

The results of this study demonstrated there was a compensation effect observed in the perceptual boundaries when the prior vowel was shortened. Thus, there is a possibility that the perceptual units of geminate/ single stops are based on a VC structure. These results coincide with results reported by Ofuka et al. (2005).

However, the results of this study suggest that further investigation is necessary before these conclusions can be extended further. Particularly, the effects of segment duration including the preceding mora, or the preceding consonant need to be further investigated.

Finally, it was determined that while Korean learners perceive the geminate/ singleton stop as a part of a VC unit the same as native speakers, the perceptual boundary is perceived differently. This difference is possibly related to the extent of temporary compensation. Further investigation is necessary to determine if this temporary compensation is a unique durational characteristic of Japanese mora timing. Related to this is Kato et al.'s (1997, 1999) hypothesis which predicts that the segmental duration is fundamental to the temporal unit controlling Japanese timing. If this is the case, so, the possibility remains that the temporal unit of the Korean affects how Korean learners of Japanese distinguish geminate and singleton stops. This needs verification in future research.

References

Fukui, S. (1978). Perception for the Japanese stop consonants with reduced and extended durations. *Onsei Gakkai Kaiho*, 159, 9-12 (in Japanese).

Kato, H., Tsuzaki, M., & Sagisaka, Y. (1997). Acceptability for temporal modification of consecutive segments in isolated words. *Journal of the Acoustical Society of America*, 101, 2311-2322.

Kato, H., Tsuzaki, M., & Sagisaka, Y. (1999). A modeling of the objective evaluation of durational rules based on auditory perceptual characteristics. In J.J.Ohala, Y.Hasegawa, M.Ohala, D.Granville, & A.C.Bailey (Eds.), *Proceedings of the 14th International Congress of Phonetic Sciences* (pp. 1835-1838). San Francisco: Regents of the University of California.

Ofuka, E., Mori, Y., & Kiritani, S. (2005). Perception of a Japanese Geminate Stop: The Effect of the Duration of the Preceding/Following Vowel. *Journal of the Phonetic Society of Japan,* 9(2), 59-65 (in Japanese).

Processing lexical tone in third-tone sandhi

Shari R. Speer* & Lei Xu#

*#Linguistics, The Ohio State University, Columbus, Ohio, USA
*speer@ling.osu.edu
#xulei0368@yahoo.com

It is not clear how prosodic structure works with segmental factors in lexical representation and access. Lexical tone, such as that found in Mandarin Chinese, is an excellent example of a prosodic, yet lexical feature that is actively involved in word recognition. Previous studies have suggested separate representation of lexical tone and segments in Chinese. For example, Cutler and Chen (1997) argued that tonal information becomes available later than segmental information during word recognition, and Ye and Connine (1999) proposed a "toneme level", operating separately from the "phoneme level" in the lexicon.

We present a series of experiments examining listeners' processing of lexical tone in relation to segmental information during word recognition in Putonghua (Standard Mandarin). In each experiment, lexical ambiguity was created through the operation of third-tone sandhi (an optional but highly productive phonological rule that changes the pronunciation of the first Tone3 word in a Tone3+Tone3 sequence to Tone2. In this way, a spoken Tone2 morpheme that has a Tone3 counterpart in the language can be intended/inerpreted as either the 'real' Tone2 or the sandhi Tone2 (Tone2s) morpheme, which is underlyingly a Tone3). Example sentence materials and visual targets are shown below.

Ambiguous Sentence, Tone2 Tone3
"jin1 nian2 hai3 bian1 **yu2 hen3** duo1"
this year sea side **fish** very much
this year near the seashore (there is) a lot of **fish**

Ambiguous Sentence, Tone2s Tone3
"jin1 nian2 hai3 bian1 **yu2s hen3** duo1"
this year sea side **rain** very much
this year near the seashore (there is) a lot of **rain**

Visual targets

Tone2 associate: shi2 pin3 食品 (food)
Tone3 associate: qing2 tian1 晴天 (sunshine)
Irrelevant: ling2 hun2 灵魂 (spirit)

A cross-modal semantic priming experiment compared the processing of tonally ambiguous and unambiguous words in spoken sentence context. Listeners heard sentence fragments that were truncated after a sandhi sequence Tone2s-3, or after a Tone3, or Tone2 word (Tone2s syllables show small but consistent phonetic differences from Tone2 syllables, see Peng, 2000), and immediately at fragment offset named visually presented Chinese character associates of the Tone2 meaning, the Tone3 meaning, or an irrelevant meaning. If Tone2 words are unambiguous without a following Tone3 context to trigger the operation of third tone sandhi, only the Tone2s-3 condition should have primed associates of both the Tone2 and Tone3 meanings. However, results showed that both forms of Tone2 were ambiguous. That is, naming times for visual semantic associates of both the Tone2 and Tone3 meanings were shorter than for irrelevant targets following a spoken Tone2 or Tone2s-3 prime. In contrast, naming times for Tone3 items were shorter for visual associates of the Tone3 meaning than for associates of the Tone2 meaning.

The cross-modal naming results suggested an integrated lexical representation of tone and segment, because morphemes pronounced with Tone2 primed semantic associates of both Tone2 and Tone3 meanings, regardless of the presence or absence of the 'trigger' for the operation of third tone sandhi. However, because we tested only Tone2 syllables that had Tone3 counterparts, it is also possible that auditory Tone2 itself is ambiguous, but still processed independently of segmental information about a syllable or morpheme. Cross-modal lexical decision experiments using words in isolation explored this possibility. Listeners heard four types of auditory words: Tone2 with Tone3 counterpart, Tone2 without Tone3 counterpart, Tone3 with Tone2 counterpart, or Tone3 without Tone3 counterpart. They made "word" / "non-word" judgments for immediately following visual characters of four types: identical word, different-only-in-tone, irrelevant and non-word. Results showed longer processing time overall for items with Tone2 and Tone3 counterparts as compared with Tone2 and Tone3 items that had no counterpart. The difference in processing time was taken to mean that the tonal information was processed in a integrated way with segmental information, and not separately at a "tonal level", which would have resulted in similar processing time for both types of Tone2. However, surface Tone2 words did not prime words with identical segments that differed only in that they had Tone3. Thus we did not find direct evidence that the Tone2 pronunciation of a CV that corresponds to both a Tone2 and a Tone3 morpheme led to lexical access of both forms. This difference between experiments might be due to the presence/absence of sentence context, or task differences between naming, which taps early lexical processing and emphasizes pronunciation, and lexical decision, which emphasizes a later, post-access judgment.

To explore this difference and examine the time-course of the resolution of lexical ambiguity due to third tone sandhi, we tracked listeners' eye-movements during a word-monitoring task. On each trial, listeners saw 4 characters: the Tone2 word, the Tone3 word, a word with the same CV but a different Chinese tone, and an irrelevant word (different CV and tone). They heard a full sentence that began with an ambiguous sentence fragment from the previous experiments, but ended with a disambiguating sequence. Spoken sentences contained temporarily ambiguous Tone2 followed by a tone other than Tone3, an ambiguous Tone2sTone3 sequence, an ambiguous Tone2Tone3 sequence or an unambiguous Tone3 followed by a tone other than Tone3. They were asked to move a cursor to the character corresponding to a word they heard in the sentence, and to click when the sentence was complete. Thus the task involved implicit syllable monitoring and tone-detection – eye movements showed participants' ongoing disambiguation process for third tone sandhi in sentence context. A comparison of results for the temporarily ambiguous and unambiguous sequences showed very early use of following tonal context to resolve the lexical ambiguity generated by a CV pronounced with Tone2. Results for the two ambiguous sequences showed a remarkable pattern: when listeners heard a Tone2Tone3 sequence, they made early looks to the character for sandhi Tone2, and when they heard a Tone2sTone3 sequence, they made early looks to the character for 'real' Tone2. Subsequent looks were to both character targets, showing awareness of ambiguity, and final looks were to the correct character as sentence context resolved the ambiguity. We interpret this pattern as sensitivity to the fine-grained phonetic differences between Tone2 and Tone2s, as the ambiguous stimuli differed only in this variable. We discuss a modified exemplar model of lexical access and representation where tone and segment are integrated, and suggest that phonetic detail may influence early aspects of tone recognition.

Acknowledgements

We thank Mary Beckman, Marjorie Chan, and the Psycholinguistics Laboratory Group for their helpful commentary during the development of this work and the Targeted Investment Fellowship Program of the Linguistics Department at The Ohio State University for financial support.

References

Cutler, A., & Chen, H.C. (1997). Lexical tone in Cantonese spoken-word processing. *Perception & Psychophysics,* 59 (2), 165-179.

Peng, S. (2000). Lexical versus 'phonological' representations of Mandarin Sandhi tones. In Michael B. Broe & Janet B. Pierrehumbert (Eds.), *Papers in Laboratory Phonology V: Acquisition and the Lexicon* (pp. 152–167). Cambridge: Cambridge University Press

Phonological learning based on interactive and mediated speech

Jane Stuart-Smith, Rachel Smith and Sophie Holmes

Department of English Language, University of Glasgow; j.stuart-smith@englang.arts.gla.ac.uk; r.smith@englang.arts.gla.ac.uk; s.holmes@englang.arts.gla.ac.uk

Introduction

Current models of language change assume that face-to-face interaction is essential for the kind of phonological learning which leads to change, and reject the possibility that such learning occurs from non-interactive speech sources such as radio and television. This assumption is challenged by recent findings showing that the broadcast media are involved in certain phonological changes. But almost no research directly investigates whether these different experiences of language cause differences in phonological learning. We report the first steps of a research programme to compare how listeners learn about an accent other than their own through face-to-face interaction versus watching a video of the same.

Background

Listening to speech from speakers in the broadcast media, who may have accents different from our own, is now an everyday occurrence for most individuals. Yet very little is known about how processes of speech perception and phonological learning differ when listening non–interactively to speech, as compared to when we participate in live interaction with others. Indeed, there are differing assumptions about the potential impact of the broadcast media on speech perception. Hay et al. (2006), working on a vowel merger in NZ English, argue for the availability of exemplars from previous media exposure, whilst Evans and Iverson (2004), who looked at the perceptual categorization of vowels in speakers of British English, with different opportunities for face-to-face interaction as a result of where they lived, seem to come to the opposite conclusion.

There are good reasons to think that interaction might give rise to special kinds of learning about the accents of others. Interactive speech may engage attention more fully, and be more memorable, than mediated speech, because it is often perceptually richer (e.g. providing three-dimensional visual as well as auditory information), and because it typically involves the collaborative pursuit of goals. Finally, interactional partners appear to align with or entrain to each other's lexical and syntactic structures, phonetic patterns, breathing patterns and speech rhythms (e.g. Pickering and Garrod, 2004).

At the same time, there is evidence from sociolinguistics that people can learn about mediated speech. Work with adolescents in Glasgow, Scotland (e.g. Stuart-Smith, 2005) shows correlations between engagement with a London-based television soap, and production of phonetic features characteristic of London. Moreover, experiments on speech perception that use non-interactive recorded speech (as most do) show that speakers can carry out many kinds of perceptual learning about such speech (e.g. Norris et al., 2003). Intriguingly, recent work testing infants' perception of a foreign-language phonological contrast suggests that live interaction with a speaker of the language gives rise to learning where equivalent exposure via video fails to (Kuhl et al., 2003). But there is no comparable research with adolescents or adults.

Experimental design and methodology

The experiment assessed how interactive vs. mediated exposure affects the way speakers of Scottish English learn about a distinction that is non-phonemic for them, but phonemic in Standard Southern British English (SSBE). The SSBE distinction investigated is between /a/ and /ɑː/, which occur in four lexical sets (Wells, 1982): TRAP (/a/), BATH, PALM and START (all /ɑː/): certain codas occur only with one vowel (e.g. /ŋ/, /ŋk/ with /a/: *hang, bank*), while other codas can occur with either vowel (e.g. /nt/: *ant, aunt*). In Scottish English, words in all four sets contain /a/ [a̠]. The experiment had 40 participants, and three phases:

1) a rhyme-judgement test, to establish baseline knowledge. It used 48 monosyllables (selected from the TRAP, BATH and PALM lexical sets) and 52 distractors. Participants judged which vowel an SSBE speaker would use in each word, categorizing it as having the vowel of *cat*, *bath*, or a

different vowel. The design grouped the experimental words equally into "more predictable"/ "less predictable," according to the extent to which /a/ or /ɑː/ might be predicted from their coda.

2) a period of either interactive or mediated exposure to SSBE. In the interactive condition, 20 participants played a word game with the SSBE-speaking experimenter, which required the experimenter to mention half of the experimental words. Each interaction was video-recorded and formed the exposure material for one of 20 participants in the mediated condition, who watched it and kept a tally of the scores as the players played the game.
3) a different randomization of the rhyme-judgement test, to examine change in participants' knowledge of the SSBE system.

The data were analysed in R using generalized linear mixed-effects modeling. Models were fitted for two dependent variables, correct responses, and consistency of responses across the rhyme-judgement tests.

Results

We expected all participants to improve at the task, and those in the interactive condition to improve more. Neither prediction was straightforwardly supported by the results, but a range of differences between the interactive and mediated conditions emerged when lexical set was taken into account. Overall, the proportion of correct responses did not increase from the first to second rhyme-judgement test, and decreased for BATH words. There was no difference in the extent to which interactive and mediated participants improved at the task; but interactive participants did make more correct responses to BATH and PALM words than mediated participants, and marginally fewer correct responses to TRAP words; there was a trend for these patterns to become more pronounced from the first rhyme-judgement test to the second.

Interactive and mediated participants also differed marginally significantly in terms of the consistency of their responses, with interactive participants more consistent than mediated participants. Responses to TRAP words were much more consistent overall than those to BATH, and responses to words not heard in the game were marginally more consistent than those to words actually heard.

Discussion

For our Scottish speakers, the hardest words to categorise were those containing /ɑː/ in SSBE (i.e. BATH, PALM). Participants who received mediated exposure performed particularly poorly on these words, and also showed less consistency in responding across the two tests than those in the interactive condition. A possible interpretation of these findings is that mediated participants underwent more disruption to their existing knowledge about the distribution of the sounds than interactive participants did, and that this disruption represents the initial stages of learning, which is made possible by the reduced cognitive load of watching a game as opposed to participating in it. Broader issues arising from the results include why participants failed to improve at the task overall, and why there was no difference in the interactive and mediated groups' improvement at the task. These will be addressed in ongoing refinements to our developing methodology.

References

Evans, B., & Iverson, P. (2004), 'Vowel normalization for accent: An investigation of best exemplar locations in northern and southern British English sentences', *JASA*, 115, 352-61.

Hay, J., Warren, P., & Drager, K. (2006), 'Factors influencing speech perception in the context of a merger-in-progress', *Journal of Phonetics,* 34 (4): 458-484.

Kuhl, P., Tsao, F-M., & Liu, H-M. (2003), 'Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning', *PNAS,* 100, 15, 9096-9101.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology,* 47, 204-238.

Pickering, M.J. & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences.* 27, 169-225.

Stuart-Smith, J. (2005). Is TV a contributory factor is accent change in adolescents? Final Report on ESRC Grant No. R000239757.

Wells, J. C. (1982). *Accents of English.* Cambridge: Cambridge University Press.

Suprasegmentals and Ethnic Stereotypes in New Zealand

Anita Szakay

Department of Linguistics, University of British Columbia, Canada; szakay@interchange.ubc.ca

Introduction

The present paper forms part of a larger research project whose aim was twofold. First, it set out to explore possible prosodic differences between the two main ethnic varieties of New Zealand English. To investigate this, a production experiment was carried out using 36 speakers. Second, it aimed to establish what suprasegmental cues listeners might use to identify Maori English and Pakeha English, the main variety used by speakers of European descent. For this reason, a perception experiment was conducted using seven different speech conditions and 107 participants. Taken together, the results from the production and the perception experiment demonstrate listeners' awareness of certain suprasegmental characteristics of the dialects. However, the results also suggest that ethnic dialect identification is often based on prosodic stereotypes.

Production Experiment

The production experiment confirmed the results of previous research demonstrating that these two varieties display distinct rhythmic qualities (e.g. Warren, 1998), as well as differing intonational patterns (e.g. Britain, 1992). Maori English is significantly more syllable-timed than Pakeha English ($p<0.0001$) (Szakay, 2006). The two varieties also differ in their use of the High Rising Terminal contour (HRT), which is a salient rise in pitch at the end of non-interrogative intonational phrases (uptalk). Overall, Maori English speakers were shown to use a significantly higher percentage of HRTs.

In addition to rhythm and intonation, the present study also investigated other prosodic characteristics of the two dialects. Statistical analyses revealed no significant differences between Maori English and Pakeha English with regard to *speech rate*, *number of pauses*, *minimum pitch* and the *standard deviation of pitch*.

Perception Experiment

The perception experiment used seven different speech conditions to isolate the precise features that listeners might attend to when identifying speaker ethnicity. Each of these retained different prosodic information in the speech signal, thus listeners were asked to perform a forced choice dialect identification task in the following conditions:

(1) duration normalised rhythm only at mean pitch across speakers
(2) rhythm only at mean pitch across speakers
(3) rhythm only at speaker's own mean pitch
(4) intonation only
(5) rhythm and intonation together
(6) low-pass filtered at 400Hz
(7) normal speech

Creating the Speech Conditions

To create the rhythm only conditions, all segmental information was eliminated and the intonation was flattened. A Praat script was written that replaced each consonant and each pause by silence, while vowels were replaced by a tone complex. The tone complex was the sum of a number of cosine waves with equidistant frequencies at a sampling frequency of 8000 Hz. In Conditions One and Two, it was created at the mean pitch across all speakers according to gender, which was 118 Hz for male speakers and 188 Hz for female speakers. In Condition Three the tone complex was created at the speaker's own mean pitch. For Condition One duration normalisation was also performed, which speeds up slow speakers and slows down fast speakers, ensuring that listeners could only tune into rhythm and not speech rate. A scaling factor was defined for each speaker by dividing the mean vowel duration across all speakers by the mean vowel duration of the individual speaker. Then, the duration of each actual vowel in the passage was multiplied by the scaling factor. The same procedure was then carried out for all consonants and all pauses.

To create the intonation only condition, first a Pitch Tier was extracted from the original sound file, then it was degraded to a Point Process, which in turn was converted into a Sound Object by humming it. Praat's humming procedure involves running the sound through a sequence of second-order filters that represent five formants. In the hummed version the pitch contour is linearly interpolated between two adjacent points, even across original pauses. This has the undesirable effect that the end of an intonation phrase cannot be recognised. To avoid this, each segment originally marked as a pause in the textgrid was replaced by silence in the hummed version. This allowed HRTs to be recognised, making it possible to investigate whether participants use HRTs as a cue in this condition.

For the rhythm plus intonation condition a hummed version of each sound file was created. Each original vocalic segment was then replaced by the corresponding hummed segment, while consonants and pauses were replaced by silence. This condition differs from the rhythm only conditions in that all vowels retain their original pitch movements instead of being replaced by a monotonous tone complex.

Low-pass filtering at 400Hz was also carried out using Praat with a smoothing of 50Hz. As the filtering produces a muffled sound, the amplitude was multiplied by four to increase loudness.

In the normal speech condition the speech signal was unaltered, however, two non-standard syntactic structures were removed to ensure that listeners would strictly base their judgments on phonetic cues.

Results

Logistic regression analyses showed that both speaker rhythm and the percentage of HRTs used by the speaker facilitate accurate dialect identification in New Zealand. However, other prosodic cues achieve the opposite effect. Although the production results indicated no significant differences for speech rate, number of pauses, minimum pitch and the standard deviation of pitch, Maori English is perceived as significantly *slower*, *more hesitant*, *more monotonous* and *lower pitched* than Pakeha English. That listeners' expectations are not corroborated by evidence from speech production suggests that they are based on ethnic stereotypes.

Summary

The results of this study have indicated that there is a divergence between the production and the perception of some prosodic cues. This mismatch between production and perception has a hindering effect on accuracy in ethnic dialect identification. All in all, Maori English speech seems to be perceived as slow, hesitant, monotonous and low-pitched. Those speakers who do not conform with these expectations are often misidentified as Pakeha.

This study pioneered innovative methodological techniques to isolate the precise suprasegmental features that listeners use to identify the ethnicity of a speaker, which allowed the recognition of some of the current stereotypes New Zealanders hold about Maori and Pakeha speech.

Acknowledgements

Special thanks to Jen Hay, Alex D'Arcy, Margaret Maclagan and Paul Warren for their useful comments. The financial help provided by the University of Canterbury, New Zealand is also gratefully acknowledged.

References

Britain, D. (1992). The Use of High Rising Terminals in New Zealand English. *Language Variation and Change,* 4(1), 77-104.

Szakay, A. (2006). Rhythm and pitch as markers of ethnicity in New Zealand English. In P. Warren and C. Watson, (Eds.) *Proceedings of the 11th Australasian International Conference on Speech Science & Technology* (pp. 421-426). University of Auckland.

Warren, P. (1998). Timing Patterns in New Zealand English Rhythm. *Te Reo,* 41, 80-93.

An EPG study of palatal consonants in Arrernte.

Marija Tabain

Linguistics Program, La Trobe University, Melbourne, Australia.
m.tabain@latrobe.edu.au

This study presents EPG (electropalatography) data on the palatal stop "ty", nasal "ny" and lateral "ly" in Arrernte, a language of Central Australia.

The present study is motivated by Recasens and Espinosa's (2006) study of the palatal stop, nasal and lateral in Catalan, which found certain differences between the articulation of the stop and the articulation of the lateral and nasal. However, whereas in Catalan the palatal stop is an allophone of the velar stop, in Arrernte the palatal stop is a phoneme in its own right, contrasting both with a velar stop and with other coronals and the bilabial. The Arrernte palatal consonants are characterized as laminal phonologically, and the other coronals in the language are lamino-dental, apico-alveolar, and apico-postalveolar.

Recasens and Espinosa found that the palatal stop in Catalan exhibited certain articulatory characteristics of dorsal consonants (e.g. a high degree of coarticulation), in contrast to the palatal nasal and lateral. They also found a more anterior place of articulation for the lateral and nasal than for the stop. There was also an important amount of individual speaker variation from the five male speakers used in their study.

In the present study, two female speakers of Arrernte were recorded reading a list of real words in their language. Data from the Reading EPG were analysed using the following traditional EPG measures: Duration (duration of closure); COG (centre of gravity); Sum (total electrodes contacted); AI (anteriority index); CI (centrality index); and DI (dorsopalatal index).

Results show that both Arrernte speakers had a shorter duration for the palatal stop than for the nasal and lateral. More surprisingly, however, both speakers showed the least central contact for the palatal stop (CI). It is probable that this pattern reflects the build-up of intra-oral pressure behind the palatal stop constriction; however, it is also possible that, in contrast to the Catalan palatal stop, it reflects a more anterior closure target.

The remaining contact measures showed some small differences between the two Arrernte speakers. The younger speaker showed greater overall contact (Sum), and the most contact in the posterior regions of the palate (DI), for the nasal palatal. The lateral showed the least contact in the posterior regions of the palate, with the stop in between. The older speaker may have had a slightly more posterior articulation for the lateral than for the stop or nasal (AI), although this may be a reflection of a more variable contact pattern for the lateral.

These results for Arrernte are slightly different to the results for Catalan, in that there is no evidence for more anterior contact for the lateral and nasal; rather, the Arrernte data suggest fewer central contacts for the stop. Since Arrernte has so few vowels (most analyses posit only two, /ɐ/ and /ə/; another two possible vowels, /i/ and /u/, are extremely infrequent), it is difficult to know what coarticulatory differences may exist between the different manners of articulation in comparison with Catalan. However, further research will look at position in the word (initial, medial or final) as a possible factor in palatal consonant articulation.

To conclude, the present results suggest that homorganic stop and nasal palatals do not exhibit the same EPG contact patterns. It is likely that aerodynamic factors reduce the amount of contact observed for the stop in contrast to the nasal.

Acknowledgements

This work was supported by the Australian Research Council. I would like to thank Kristine Rickard for her work on labeling the database.

References

Recasens, D., & Espinosa, A. (2006). Articulatory, positional and contextual characteristics of palatal consonants: evidence from Majorcan Catalan. *Journal of Phonetics*, 34, 295-318.

Here, [ðɜː] and everywhere: lexical access, usage and the retrieval of non-linguistic information in language production

Phillip Tipton

Department of Language and Linguistics, University of Essex, UK; ptipto@essex.ac.uk
School of Languages, University of Salford, UK; P.Tipton@salford.ac.uk

Psycholinguistic models of language production have very rarely, if at all, been considered in studies of sociolinguistic variation despite the raw data being, in most cases, derived from instances of speech production (though see Preston, 2004, for an explicitly psycholinguistic take on sociolinguistic variation and Mendoza-Denton et al., 2003, for an explicit focus of production). Moreover, phonetic intra-speaker variation is rarely explicitly addressed in the models of speech production proposed. The different models of lexical access implicitly assume one phonetic form per lexical item (e.g. Levelt, 1989), but traditional variationist studies have made it plain that more than phonetic form is accessed for those instances where variation is present on an intra-speaker level. Intra-speaker variation may be found to correlate with so-called 'internal' constraints such as preceding and following phonological environments. As inadequate as this correlation might be in terms of explanation, it has even less to recommend it when it comes to accounting for different manifestations of the same lexical item. Speakers who display such phenomena in their language production are often, at least in the sociolinguistic literature, apt to be treated as 'outliers' or anomalous exceptions to the prevailing variety of the speech community.

Recent perceptual studies of the indexical capabilities of the lexicon have pointed to usage as the key to the development, juxtaposition and entrenchment of phonetic and social information in the representations of the speaker-hearer (Hay, Nolan and Drager, 2006; Hay, Warren and Drager, 2006). The approach described in this poster attempts to build on this perceptual work by offering a more production-orientated framework which, it will be argued, can fruitfully be employed in analyses of sociolinguistic micro-variation. To this end, the focus will be on language processing in the individual speaker.

In the current study, two speakers have been analysed in terms of their realisation of the so-called NURSE~SQUARE merger (Wells, 1982) which is found in, *inter alia*, parts of the north-west of England, in the United Kingdom. In general, the vowel alternation in a situation of complete contrast would be [ɜː]~[ɛː] whereas the vowel in speakers who exhibit a merged production is [ɜː]. A number of different types of data were collected from these speakers: informal conversation, word list productions and the production relating to the completion of a map task. The tokens were then analysed auditorily and classification was made on the basis of their status as either merged or unmerged The two female speakers, Bev and Mandy, aged nineteen and seventeen respectively at the time of data collection, were born, and lived ever since, in the town of Newton-le-Willows (pop. 21000). Newton is part of the heavily urbanised south of the north-west region and was built on the earliest incarnations of the railway industry; it currently houses a socially-mixed population. It lies almost equidistant (c.30km) from both Liverpool and Manchester, the two most salient urban foci, and many residents utilise the local railway stations of Newton-le-Willows and Earlestown to commute to work in these cities. The NURSE~SQUARE merger is said to be typical of certain south Lancashire varieties. The Survey of English Dialects (Orton and Dieth, 1962) examined a few locations in the vicinity of Newton. Two particularly relevant locations, Bickerstaffe and Halewood are shown to be fully merged, though these are some distance from Newton itself. Newton sits between the two larger towns of St Helens (pop. 176000) and Warrington (pop. 191000), and is administered from the former. The evidence, both from the Bickerstaffe SED data and recent small-scale studies of St Helens suggest that the merger is still robust in this region. More limited evidence exists for Warrington, but this seems to indicate that, for younger speakers at least, the linguistic situation is the complete opposite of the one extant in St Helens, that is, one of complete vocalic contrast.

This poster will deal primarily with the lexical item {there}, which is not only extremely frequent in independent corpora of English, but is also to be found extensively in the language production of the Newton speakers. {there} is, moreover, highly variable in its production, and it will be shown how the phonetic manifestation of the item can correlate very strongly with discourse topic which often, in the current case of Newton-le-Willows, centres around the assumed internal social divisions within Newton itself. The following examples come from Bev in an attempt to vocalise the speech of hypothesised Newton residents

(1) "Yeah, we live in Earlestown, down [ðɜː]"
(2) "And us Newtoners live [ðɛː]"

The unmerged manifestation occurs more frequently when the main topic of discussion is either the Newton 'end' of Newton-le-Willows or some more global concern. A merged phonetic variant, however, is dominant when the speakers are talking of the Earlestown 'end' which is the more economically-disadvantaged area and which also contains a good deal of social housing.

It will be argued that this type of variation – where correlations can be observed between social and linguistic information - is non-trivial and must be accounted for in a more integrated fashion than can be achieved by sole recourse to either traditional models of lexical access or the speculatory 'attitudinal' explanations which are often implicit in the variationist literature. As a result it will further be argued that Preston's notion of a 'sociocultural selection device' (2004) is both insufficiently specific in its architecture and, at the same time, too reliant on an albeit weak version of the competing grammars approach to variability. Instead, it will be proposed that social information has properties similar to the kind of material that is found in the pre-phonological stages of lexicalization models, thus rendering surface sociolinguistically-patterned phonetic variation a product of the retrieval of separate lemmas. An interactive view of lexicalization will be assumed and the implications for usage-based approaches to linguistic variation will be discussed.

References

Hay, J., Nolan, A., & Drager, K. (2006). From Fush to Feesh: Exemplar Priming in Speech Perception. *The Linguistic Review*, 23, 351-379.

Hay, J., Warren, P., & Drager, K. (2006). Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics*, 34, 458-484

Levelt, W. (1989). *Speaking*. Cambridge, MA: MIT Press.

Mendoza-Denton, N., Hay, J., & Jannedy, S. (2003). Probabilistic Sociolinguistics: Beyond Variable Rules. In R. Bod, J. Hay & S. Jannedy (Eds.), *Probabilistic Linguistics* (pp. 97-138). Cambridge, MA: MIT Press.

Orton, H., & E. Dieth, E (Eds.). (1962). *The Survey of English Dialects.* Leeds: Arnold.

Preston, D. (2004). Three Kinds of Sociolinguistics: A Psycholinguistic Perspective. In C. Fought (Ed.) *Sociolinguistic Variation: Critical Reflections* (pp. 140-58). Oxford: Oxford University Press.

Wells, J. (1982). *Accents of English.* Cambridge: Cambridge University Press.

Phonetic Details and Lexical Idiosyncrasies in Vowel Categorization in Taiwanese

Jane Tsay[1] & Yunwei Lee[2]

Institute of Linguistics, National Chung Cheng University, Taiwan;
[1]lngtsay@ccu.edu.tw; [2]astywl@ccu.edu.tw

This paper aims to demonstrate how phonetic details and lexical idiosyncrasies influence the categorization of vowels in a sound system, using natural speech from Taiwanese (also called Taiwan Southern Min). It is widely accepted in the literature that this language has six non-nasal monophthongs, i.e. /i/, /e/, /a/, /u/, /ɔ/, and /o/ (e.g. Yang, 1991; Zhang, 1993; Chung, 1996). However, to our knowledge, there have not been acoustic studies reported to support the vowel categorization in this language. In this study, F1 and F2 of vowels in open syllables were extracted from natural speech (410 minutes) of a 30-year-old female. We have the following interesting findings. First, comparing mean F1 and F2 values of /o/ and /ɔ/ in Taiwanese with those in English (as reported in Bardlow, 1995), the values in Taiwanese were both much higher (Table 1).

Table 1: Mean F1 and F2 of /ɔ/ and /o/ in Taiwanese and English

Vowel	Language	Mean F1 (Hz)	Mean F2 (Hz)
/ɔ/	English	620	1033
	Taiwanese	875	1609
/o/	English	482	1160
	Taiwanese	739	1633

That means Taiwanese /ɔ/ and /o/ are lower and more front than English /ɔ/ and /o/, respectively. Secondly, there was a visible overlap between /o/ and /ɔ/ in the vowel space, larger than that of any other two neighboring vowel categories (Figure 1), although still statistically distinct ($p<0.05$).

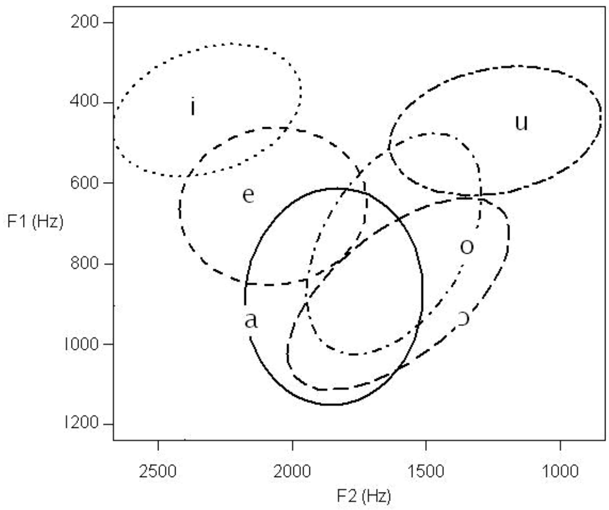


Figure 1: Vowel distribution of Taiwanese

After examining the phonetic details of /ɔ/ and /o/ in Taiwanese, we found that the distribution of one lexical item /ho/ 'good' is separated from most of the other words with the /o/ vowel, being located in a lower area away from the other words (Figure 2).

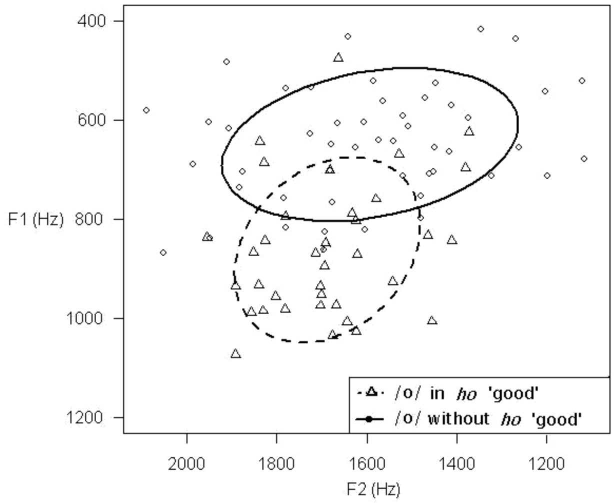


Figure 2: /ho/ 'good' vs. /o/ without /ho/

Furthermore, the vowel distributions of the minimal pair /ho/ 'good' and /hɔ/ 'give' showed that the two vowels were not significantly different (p>0.1) and might be merging (Figure 3).

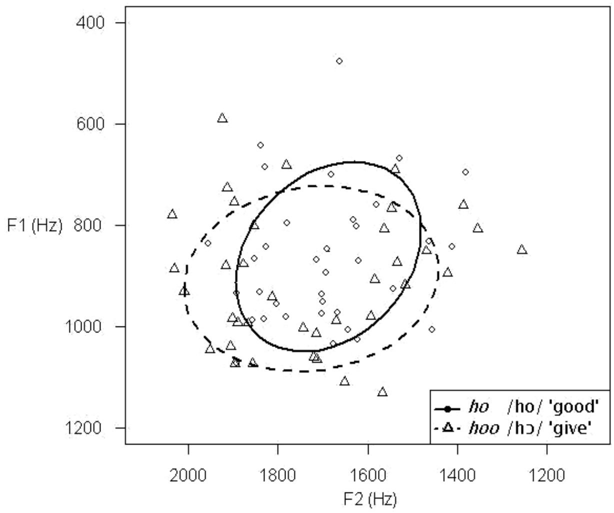


Figure 3: /ho/ 'good' vs. /hɔ/ 'give'

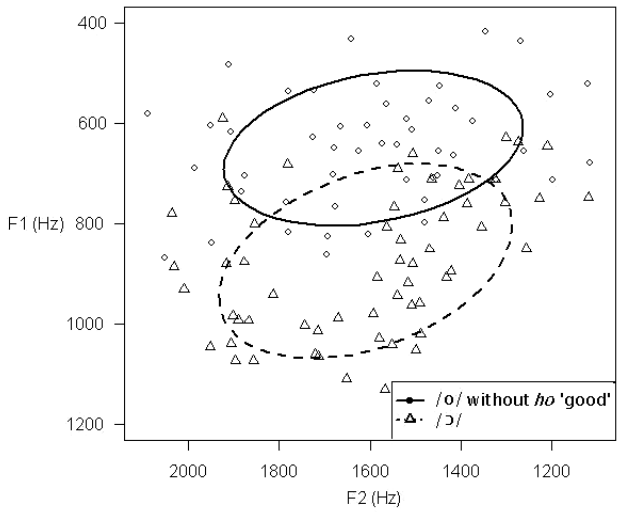


Figure 4: /o/ without /ho/ 'good' vs. /ɔ/

After excluding the tokens of /ho/ 'good', /o/ and /ɔ/ became more separated in the vowel space (Figure 4). This suggests that it is the idiosyncratic /ho/ 'good' that causes the overlapping phenomenon between /o/ and /ɔ/ vowel categories. Hence, whether there is a sound change in progress in /o/ in Taiwanese is worth further investigation.

References

Bradlow, Ann R. (1995). A comparative acoustic study of English and Spanish vowels. *Journal of the Acoustical Society of America*, 97(3), 1916-1924.

Chung, Raung-fu (1996). *The Segmental Phonology of Southern Min in Taiwan.* Taipei: Crane.

Yang, Hsiu-fang (1991). *Taiwan Minnanyu Yufagao [Taiwanese Southern Min Grammar].* Taipei: Da-an.

Zhang, Zhen-xing (1993). Taiwan Minnan Fangyien Jilue [A Note on Taiwan Minnan Dialects]. Taipei: Wen Shi Zhe.

Ongoing merger of lexical tonal classes in Masan/Changwon Korean: Evidence from lexical and postlexical/intonational phenomena

Akira Utsugi

JSPS Postdoctoral Fellow for Research Abroad/University of Edinburgh, UK; autsugi@ling.ed.ac.uk

Introduction

This presentation reports the merger of lexical tonal classes in Masan/Changwon Korean (MCK) so as to contribute empirically to general issues on tonal phenomena, such as lexical representation, diachronic change, and their interaction, as well as to language-specific issues in Korean prosody. MCK has sparsely specified tones, some of which have a distinctive function, similar to Tokyo Japanese (see Utsugi, 2007 for details of the prosody of MCK). The present study focuses on Low-High (LH) patterns in bisyllabic words. There are two lexical tonal classes, called 'Final' and 'Medial-Double', which show LH patterns at the surface. The two classes have diachronically different origins, and are clearly distinguished even in citation forms in an older generation. However, they seem to be in the process of merging in a younger generation. This study shows how they are merging in lexical and postlexical aspects.

LH in the older generation

According to Gim (1970), pitch patterns of bisyllabic Final and Medial-Double tonal classes are different in terms of a tonal register in the older generation; Final has a Mid-High (MH) pattern, whereas Medial-Double has a LM pattern. In addition to the difference in citation forms, there are differences in other forms.

When the two classes are followed by postpositional particles, they show a difference in the particles; low pitch after Final and high (or mid) pitch after Medial-Double (e.g., Gim, 1970).

Compounds also show a difference. For example, in the compounds consisting of two bisyllabic morphemes, when the first component is Final and the second is also Final, they show a LH+HH pattern. When the first is Final and the second is Medial-Double, they show a LH+LL pattern. On the other hand, when the first component is Medial-Double, they show a LH+HL pattern irrespective of the second component (e.g., Hayata, 1974).

Sentence-level pitch (intonation) also seems to show a difference, although it has not been studied well. In a preliminary study of the younger generation by the author, two major patterns were found when focused Final or Medial-Double words were followed by post-focus words whose original tonal pattern was LHLL. In one pattern, a high plateau (or sometimes a gradual rise) was found from the final syllable in the focused word to the second syllable in the post-focus word (i.e. LH#HHLL). In the other pattern, a high plateau was found from the second syllable in the focused word to the first syllable of the post-focus word (i.e. LH#HLLL). It is assumed that, in the older generation, the former pattern corresponds to a pattern in phrases whose first word is a Final, and the latter pattern corresponds to a pattern in phrases whose first word is a Medial-Double. This assumption is based on the similarities to the patterns in the forms mentioned above; Medial-Double always shows the LHHL (or LMML according to Gim, 1970) pattern.

Ongoing merger in the younger generation: purpose, methods, and expectation

In the younger generation of MCK and some neighbouring dialects, it has been reported that the two classes do not show the difference in the tonal register in citation forms (e.g., Utsugi, 2007 for MCK; Kenstowicz, Cho, & Kim, 2008 for Busan Korean). It is assumed that this absence of distinction motivates the merger of the two classes. The present study investigated the pitch patterns in utterances other than citation forms to see the actual process of the merger.

Recordings of seven native speakers in their twenties were made. Materials were based on five words for each of the originally Final and Medial-Double classes. Based on these words, three sets of the material were created: (i) words followed by particles, (ii) compounds, and (iii) focused sentences (hereafter, ‘Particles’, ‘Compounds’, and ‘Focus’). As reviewed in the previous section, in all these types, patterns originated from Final (‘Final patterns’) and Medial-Double (‘Medial-Double patterns’) are different.

If the two classes are not merged in the younger speakers, it is expected that the originally Final words show the Final patterns and the originally Medial-Double words show the Medial-Double patterns in all the material sets. In this case, the absence of the distinction in citation forms is interpreted as neutralisation. If the two classes are completely merged, it is expected that all of the tokens show either the Final patterns or the Medial-Double patterns.

Results and discussion

Results were more complicated than the above expectations. Figure 1 shows how the originally Final and Medial classes were actually pronounced by three speakers. In the subject HYJ, most of the tokens were pronounced as the Medial-Double patterns regardless of the original tonal classes and the material sets, suggesting that the Final was almost merged into the Medial-Double (Figure 1a). In KMJ, most of the tokens in Focus and Compounds showed Medial-Double patterns, suggesting that the Final was almost merged into the Medial-Double as in HYJ (Figure 1b, the upper four bars). However, the opposite tendency was found in Particles (Figure 1b, the lowest two bars). This suggests a between-material-set difference and a between-speaker difference. In JHH, both the two patterns coexisted, but the preference for the patterns seems to differ among material sets (Figure 1c).

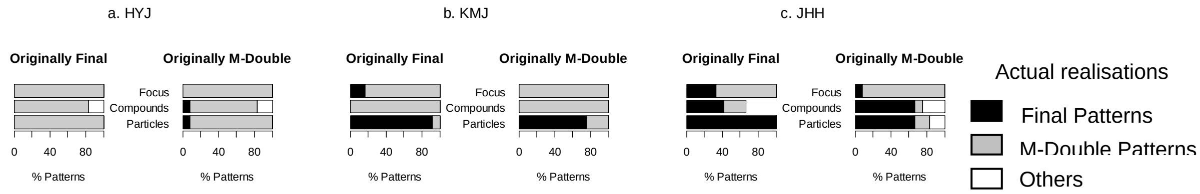


Figure 1: Results of the present research.

These results suggest that the two patterns having diachronically different origins coexist in the dialect, and even in a speaker in some cases. The preference of the patterns differs with speakers and grammatical conditions. It is unclear whether the patterns bear different intonational meanings and/or sociolinguistic indices, or they are just selected at random in the younger generation.

Utsugi (2007) proposed an intonational phonological model in which the Final pattern has H*+L lexical pitch accent, whereas the Medial-Double pattern is associated with L_{ω}+H+H+L lexical word-edge tones from the right edge of the phonological word, to account for different postlexical behaviours between the two. How can we reconcile such a model with the variation found in the present study? Does a word have two underlying representations at the same time, one of which is realised at the surface according to some conditions? Or is the surface variation produced through the phonetic and/or phonological process while keeping a single underlying representation? This question is open at this point.

References

Gim, C.-G. (1970). Kyengnam pangen-uy sengco yenkwu. *Hangul*, 145, 109-149.

Hayata, T. (1974). Accent in Korean: Synchronic and diachronic studies. *Gengo Kenkyu*, 66, 73-116.

Kenstowicz, M., Cho, H., & Kim, J. (2008). A note on contrasts, mergers, and acquisitions in Kyungsang accent. *Toronto Working Papers in Linguistic*, 28, 107-122.

Utsugi, A. (2007). *The interplay between lexical and postlexical tonal phenomena and the prosodic structure in Masan/Changwon Korean*. Workshop on Intonational Phonology: Understudied or Fieldwork Languages, International Congress of Phonetic Sciences. Saarbrücken, Germany.

Phonetic variation of /r/ in a language contact context: The case of South Tyrol Italian

Alessandro Vietti & Lorenzo Spreafico

Centre for Language Studies, Free University of Bolzano / Bozen, Italy;
Alessandro.Vietti@unibz.it, Lorenzo.Spreafico@unibz.it

Our aim is to describe the phonetic variants of /r/ in South Tyrol Italian (STI) and to link this variation to both linguistic (phonetic-phonological) and extra-linguistic (typologies of bilingual speaker) factors (Flege, 2007).

South Tyrol is a bilingual alpine region situated in Italy (since 1919) that lies on the Germanic – (Italo) Romance dialectological border. Although this area has been described as a societal bilingualism with two quite separate linguistic communities, (Bavarian) German and Italian (Mioni, 1990), nowadays the actual degree of overlapping between the two speech communities as well as the number of bilingual speakers seems to be increasing

In this sociolinguistic context, South Tyrol Italian shows a higher degree of /r/ variation if compared to the common standard-dialect Italo-Romance linguistic repertoire (a preliminary instrumental analysis of /r/-sounds in Italy is offered in Romano, in prep.). Standard Italian is usually reported to have two or three (not so neatly contextual) realizations of /r/: The alveolar trill [r], the geminate [rr], and the alveolar tap [ɾ], e.g. *raro* [ˈraːɾo] /ˈraro/ ‘rare’, *carro* [ˈkarːɾo] /ˈkarro/ ‘cart, wagon’ (Canepari, 1999). Although in South Tyrol there is not a lively presence of Italo-Romance dialects, the neighbouring Trentino-Veneto dialects exerted a strong influence on STI (due to strong immigration fluxes in the Twenties). Thus, typical Venetian /r/ variants, such as the post-alveolar (retroflex) flap [ɽ] and approximant [ɹ̣], are to be found even in STI. To complete the complex picture of /r/ phonetic sub-system, Bavarian German dialects back /r/-sounds (trill [ʀ], fricative [ʁ], and approximant [ʁ̞]) are well attested in STI. (a characterisation of /r/-sounds in Germany is offered in Wiese, 2001).

Our data are drawn from 10 speakers (3 males and 7 females) living and/or working in Bolzano (the political and administrative centre of South Tyrol). The sample is stratified according to three broad categories of bilingual speaker: “early bilingual” (roughly a speaker with two native languages acquired during primary socialization), “German native speaker” (a late bilingual that acquired Italian as a language of secondary socialization, e.g. at school or from the peer group) and “Italian native speaker” (a late bilingual that acquired German as a language of secondary socialization).

The elicitation techniques included reading tasks (e.g. word and sentence lists), visual *stimuli* and map tasks, using a modified version of the CLIPS (Corpora and Lexicon of Spoken Italian) protocol. However the present analysis is based only on reading tasks. The elicitation list contains sets of words controlled for frequency and phonological relevant variables.

In order to assess if the high degree of /r/ variation in STI is context-independent or due to allophonic variation (as reported -at least to some extent- for standard Italian), special attention was paid to the selection of words displaying different phonetic environments (especially consonant clusters), syllabic positions (e.g. in the onset: r*espinge*re ‘to push back’ or in the coda: *fo*r*te* ‘strong’) and stress patterns.

The main goal of our analysis is to provide an instrumental aided description of the different spectral-acoustic properties of /r/-sounds, with special attention to the back variants (uvular trill, fricative and approximant). Then, we seek for contextual distribution of the variants (against a free variation hypothesis).

On an extra-linguistic level we look for relationships between sets of variants (front vs. back) and the three categories of speakers, early bilinguals and late bilinguals (German and Italian native speakers). Our main goal is to look for a relation between the occurrence of back variants and late bilingualism (of German native speakers). Back variants are thus considered as markers of German nativeness.

References

Canepari, L. (1986). *Italiano standard e pronunce regionali*. Padova: CLEUP.

Flege, J. (2007). Language contact in bilingualism: Phonetic system interactions. In J. Cole & J. I. Hualde (Eds.), *Laboratory Phonology 9* (pp. 353-382). Berlin: Mouton de Gruyter.

Mioni, A. (1990). Bilinguismo intra- e intercomunitario in Alto Adige/Südtirol: considerazioni sociolinguistiche. In F. Lanthaler (Ed.), *Mehr als eine Sprache. Zu einer Sprachstrategie in Südtirol - Più di una lingua. Per un progetto linguistico in Alto Adige* (pp. 13-36). Merano: Alpha & Beta.

Romano, A. (in prep.). A contribution to the study of phonetic variation of /r/ in French and Italian linguistic domains. Proceedings of 'r-atics2': 2nd International Workshop on the Sociolinguistic, Phonetic and Phonological Characteristics of /r/ (Université Libre de Bruxelles, 5-7 Dec. 2002).

Wiese, R. (2001). The unity and variation of German /r/. In R. van Hout & H. van de Velde (Eds.), *r-atics, Sociolinguistics, Phonetic and Phonological characteristics of /r/*, Etudes et Travaux, 4, 11-26.

Compensation for Coarticulation may reflect gestural perception: Evidence from a critical examination of the effects of non-speech contexts on speech categorization

Navin Viswanathan, James S. Magnuson, & Carol A. Fowler
University of Connecticut & Haskins Laboratories

Introduction There are two prominent views regarding the "atoms" of phonology. One is that phonological units are mental categories that are modified, distorted, or destroyed during speech production (e.g. Hockett, 1955). The corresponding theory of speech perception is that listeners infer these intended mental categories from speech signals by means of an inferential process of cue matching (e.g. the general auditory viewpoint; Diehl, Lotto, & Holt, 2004). An alternate view is that vocal tract gestures in speech production, rather than being inferior surrogates of ideal mental categories, are themselves the basic units of phonology (articulatory phonology; Browman & Goldstein, 1992). The corresponding theory of speech perception is the direct realist theory (Fowler, 1986), which asserts that the acoustic signal contains lawful information about the vocal tract gestures that produced it, and listeners perceive these gestures *directly* (without inference or hypothesis testing).

Both views have appealed to results from the compensation for coarticulation (CfC) paradigm for support. In CfC, listeners' categorization of speech segments is affected by properties of the surrounding segments. According to the gestural viewpoint, CfC reflects listeners' attunement to articulatory contingencies of coarticulated speech. For instance, listeners perceive more [ga]s in a [ga]-[da] continuum when continuum members are preceded by [al] rather than [ar] because, in the production of [alga], the point of constriction during [ga] is pulled forward by the front place of constriction for [al]. According to the general auditory view, CfC reflects *spectral contrast*. In the previous example, [al] has a high F3 offset as compared to [ar] and this leads the F3 of the following stops to be heard as lower. Because a low F3 provides information for [ga] (in a [ga]-[da] continuum), listeners report more [ga]s. While in this case, the two accounts make same predictions (albeit for completely different reasons), other contexts exist where the two accounts can be dissociated (e.g. Viswanathan, Magnuson & Fowler, submitted). Determining the true origin of CfC not only paves way for a unified account of CfC but will help adjudicate between vastly different theories of phonology.

Lotto and Kluender (1998) showed that, when the precursor speech tokens are replaced by pure sine tones at the F3 offsets of [al] and [ar], listeners show CfC-like responses (more [ga] responses for the high "al" tone than the low "ar" tone). Several experiments with pure sine tones have demonstrated CfC-like effects, bringing into question phonological explanations for CfC. However, no one has tested whether the conditions required for spectral contrast are present in natural speech. When compared to natural speech formant transitions, tone analogues (i) are higher in intensity (matched to syllable rather than formant intensity), (ii) do not change in frequency (but see Lotto & Kluender, 1998) and (iii) have extremely narrow bandwidth. Our approach, through our experiments, is to examine how the effects of spectral contrast change as the non-speech sine tones are modified to be more closely matched to natural speech along the dimensions described above.

Method An 11-step series of resynthesized CV syllables varying in F3-onset frequency (in 100 Hz steps from 1800 Hz ([ga]) to 2800 Hz ([da]), with a duration of 215 ms, was created using the source-filter method. F1, F2, and F4 were the same for all continuum steps. Across conditions, the CV syllables were preceded by different precursors. **Natural speech precursors** were the syllables [al] and [ar], matched for duration (each 375 ms long) and intensity (70 dB SPL) with critical F3 offset of [al] at 2600 Hz and [ar] at 1820 Hz. T**ypical tone analogues** consisted of steady sinewave tones at the frequency of the natural syllables' F3 offsets presented at the overall intensity of the entire natural syllable (70 db SPL), which is the typical procedure in such studies (e.g., Lotto & Kluender, 1998). The **intensity-matched tones** were the same tones but presented at the intensity of the syllable's F3 (48 dB for [al] and 52 dB for [ar]). Tones tracking the entire F3 trajectories of the natural precursors, presented at F3's intensity, were used in the **intensity-matched transient tone** condition. Finally, the **filtered speech** precursors resembled the natural syllables' F3 on all three dimensions. A Hanning Band-pass filter isolated frequencies in the range of the critical F3 region of the precursor syllables (1600 Hz to 3000 Hz with a smoothing of 100 Hz). In each condition, each precursor was combined with one token from the stop continuum with a gap of 50 ms. resulting in a total of 22 tokens (2 precursor X 11 target) per condition.

110 undergraduates participated for course credit. The task was a two-alternative forced-choice: after hearing a precursor followed by one of the continuum CVs, participants pressed keys labeled "d" or "g" to indicate their identification of the stop. The experiment consisted of two blocks of trials. In the first (practice) block, only the [da] and [ga] endpoints were presented without precursors, and responses were followed by feedback. There were 12 trials with each endpoint, presented in random order. In the second block, each item from the 11-step [da]-[ga] continuum was presented in both liquid contexts eight times each resulting in 176 disyllable trials. No feedback was provided, and participants could take a break after every 44 trials.

Results Data from 100 participants (20 per condition) who scored more than 80% on the endpoint judgment task were included in the analysis. CfC was calculated by subtracting the mean percentage of "g" responses to [ar] (or, in nonspeech conditions, its analog) from the mean percentage of "g" responses to [al] (or its analog), at each step of the continuum, for each subject. These difference scores were submitted to a 5 (condition) X 11(step) mixed ANOVA. The effect of condition was significant (F (4, 95) = 9.19, p <.001, $\eta_p^2 = 0.28$), indicating that the amount of compensation varied across conditions. The effect of step (F (10, 950) = 24.38, p <.001, $\eta_p^2 = 0.20$) was also significant, because compensation varied across the continuum. Post-hoc tests showed that the intensity-matched transient tones (F (1, 38) = 9.92, p <.003, $\eta^2 = 0.21$) and the filtered speech (F (1, 38) = 16.22, p <.001, $\eta^2 = 0.30$), the two conditions most closely resembling speech showed significantly less compensation than the speech condition.

Conclusion Appeal to spectral contrast as the basis of CfC requires that the conditions for spectral contrast be present in natural speech whenever CfC is observed. Our data indicate that, as the non-speech conditions are made increasingly like speech in their acoustic properties, contrast effects diminish. Presenting only the critical region of a speech syllable that is thought to be the locus of spectral contrast is not sufficient to observe CfC. This weakens the general auditory claim that the basis for CfC is purely spectral contrast. Moreover, other studies show that CfC can occur in the absence of spectral contrast (e.g. Mitterer, 2006) or even in the opposite direction of that predicted by spectral contrast (Viswanathan et al., submitted). These results and the current findings are not compatible with spectral contrast (but are not incompatible with other general cognitive accounts of speech perception, e.g., Nusbaum & Magnuson, 1997). They support the account of CfC that follows from gestural phonology and direct realism that CfC reflects listeners' perception and attunement to vocal tract gestures, and therefore CfC results from information specifying place of articulation rather than from low-level auditory properties. Of course, uncovering the nature of information in the speech signal that drives CfC would greatly enhance our understanding of the basic units of phonology, and our research program includes efforts to isolate that information.

References

Browman, C.P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, 49, 155-180.

Diehl, R.L., Lotto, A.J., & Holt, L.L. (2004). Speech perception. *Annual Review of Psychology*, 55, 149-179.

Fowler, C. A. (1986). An event approach to a theory of speech perception from a direct-realist perspective. *Journal of Phonetics,* 14, 3-28.

Hockett, C. F. (1955). *A Manual of Phonology*. Baltimore: Waverley Press.

Lotto, A. J., & Kluender, K. R. (1998). General contrast effects of speech perception: Effect of preceding liquid on stop consonant identification. *Perception & Psychophysics*, 60, 602-619.

Mitterer, H. (2006). On the causes of compensation for coarticulation: Evidence for phonological mediation. *Perception & Psychophysics,* 68, 1227-1240.

Nusbaum, H.C., & Magnuson, J.S. (1997). Talker normalization: Phonetic constancy as a cognitive process. In K. Johnson & J. W.Mullennix (Eds.), *Talker Variability in Speech Processing* (pp. 109 – 132). New York: Academic Press.

Viswanathan, N., Magnuson, J.S., & Fowler, C.A. (submitted) Compensation for coarticulation: Disentangling auditory and gestural theories of perception of coarticulatory effects in speech.

Detailed Phonetic Memory for Multi-Word and Part-Word Sequences

Travis Wade & Bernd Möbius

Institute for Natural Language Processing, University of Stuttgart, Germany;
{travis.wade, bernd.moebius}@ims.uni-stuttgart.de

Humans recognize previously heard spoken words better when the words are repeated in the same voice and at the same speaking rate than when they differ in these or other dimensions (e.g., Goldinger, 1996). Such findings have usually been taken as evidence that perceived instances of either words or sublexical units are stored in a detailed form in memory, and that collections of these episodic memory traces comprise or are somehow linked to mental lexical representations. A different possibility is that, while detailed information about the speech signal is remembered, this information is not explicitly segmented or organized into discrete, traditionally assumed units such as words. Instead, episodic memories may involve sequences of connected speech that vary in length depending on the nature of the listening situation, perhaps comprising entire phrases or utterances. Such a possibility is more in line with the temporal and context-dependent nature of speech perception and production than traditionally defined exemplar models, since it predicts that more local information is always stored and (potentially) considered as part of a larger context. Indeed, there is evidence from corpus studies of pronunciation (Binnenpoorte, Cucchiarini, Boves, & Strik, 2005), phonological analysis (Bybee, 2002 for a review), and word-monitoring experiments (Sosa and MacFarlane, 2002) that at least very frequent multi-word expressions are in some sense accessed as "units" in the same way that words are usually assumed to be. However, it is unknown whether variable-length stretches of connected speech in general exhibit properties of exemplar-based storage or perception, since previous studies on memory have considered only isolated word productions.

In this study, word identification and recognition memory accuracy were measured for sequences of words extracted from a large multi-speaker corpus of German sentence productions (Kohler, 1996). Sequences varied in their length (0.5-3 words, in 0.5-word steps), onset phase with respect to word onset (beginning either at a word onset or halfway between a word onset and offset), and speaker (target sequences were repeated with the same or a different voice). Sequences were extracted at random from the corpus, without considering word or sequence frequency or probability or syntactic constituent boundaries. In Experiment 1, listeners (n=19) heard 384 sequences (selected separately from the corpus for each listener to maximize coverage) in two blocks and identified the words they perceived. Blocks contained the same 192 word (or part-word) sequences extracted from the same sentence contexts; half of the words in the second block were identical repetitions of first-block stimuli and half were produced by a different speaker. Experiment 2 also involved two blocks. The first block was identical to that of Experiment 1; in the second block, listeners (n=16) categorized sequences as new or previously encountered. Half of repeated sequences (one fourth of the trials in the second block) were identical repetitions, and half involved a new speaker.

Word identification results (Experiment 1) were consistent with previous findings concerning the effects of local acoustic context on phonetic perception, and the influence of lexical information on speech segmentation (e.g., Mattys, White, and Melhorn, 2005). By several measures of accuracy and consistency, listeners were better at identifying words in longer sequences, approaching ceiling performance for sequences longer than about 2.5 words. Sequence onset phase was also important, with robustly better identification where stimulus onsets and/or offsets coincided with word onsets/offsets. Responses were more consistent across repetitions of a sequence by the same speaker than repetitions by different speakers, especially at shorter sequence lengths.

Recognition memory results (Experiment 2) differed from these patterns in several critical ways. Recognition accuracy increased linearly with sequence length over the range of stimuli considered, even after identification accuracy reached ceiling performance. There was no effect of sequence onset phase on memory, and there was a significant difference in the size of the benefit of stimulus and word onset/offset coincidence between identification and memory tasks (no benefit in the memory task), indicating that lexical cues which aided segmentation of sequences did not affect the encoding of these same sequences in memory.

Both overall and within length and phase conditions, repeated sequences were more likely to be recognized as previously heard if at least one word was identified correctly during the first block than if no words were correctly identified. However, better-than-chance memory performance was seen even for sequences where no words were correctly identified. Moreover, if completely misidentified (no words correct) sequences were discarded, the likelihood of correct recognition of sequences of a given length was not related to the number of words that were originally correctly identified. This indicates that both segmentation/identification and memory were limited by auditory/perceptual factors for the most difficult stimuli; these stimuli tended to be the shortest in absolute (time) length, both overall and when considered within word length conditions. However, there was no direct evidence that words were remembered as discrete units, or that memory was directly mediated by segmentation.

Recognition memory was better for sequences repeated with the same voice than for sequences repeated with a different voice. This was true whether sequences that were completely misidentified in the first subtest were discarded or included in the analysis. The size of this same speaker recognition memory benefit tended (n.s.) to increase with sequence length, even at lengths where word identification was at or near ceiling performance. In fact, over the range of sequence lengths considered, the same speaker benefit in memory correlated inversely with the same speaker benefit in identification consistency.

In summary, a same-speaker recognition memory advantage was seen for sequences of connected speech ranging from 0.5 to 3 words in length. Memory was generally better for longer sequences, but neither overall performance nor the size of the same-speaker benefit seemed to be influenced by parsing at a lexical level, as measured by word identification. These results are consistent with a model of perception and memory in which detailed episodic storage occurs for acoustic sequences that are of variable size and composition, potentially corresponding to multiple words or phrases and not necessarily coinciding with words as discrete units. It seems further possible that processes involving various levels of a linguistic hierarchy (both phonetic and lexical, for example) may reference, or result implicitly from the structure of, the same set of multi-word episodes at multiple time scales.

References

Binnenpoorte, D., Cucchiarini, C., Boves, L., & Strik, H. (2005). Multiword expressions in spoken language: An exploratory study on pronunciation variation. *Computer Speech and Language*, 19, 433-449.

Bybee, J. (2002). Phonological evidence for exemplar storage of multiword sequences. *Studies in Second Language Acquisition*, 24, 215-221

Goldinger, S.D. (1996). Words and voices: episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 1166-1183.

Kohler, K.J. (1996). Labelled data bank of spoken standard German; the Kiel Corpus of Read/Spontaneous Speech. In *Proceedings of ICLSP 1996, Philadelphia*.

Mattys, S.L., White, L., & Melhorn, J. (2005). Integration of multiple speech segmentation cues: a hierarchical framework. *Journal of Experimental Psychology: General*, 134, 477-500.

Sosa, A.V., & MacFarlane, J (2002). Evidence for frequency-based constituents in the mental lexicon: collocations involving the word *of. Brain and Language*, 83, 227-236.

Phonetic information affects grammaticality judgements

Abby Walker

Department of Linguistics, University of Canterbury, New Zealand; ajw129@gmail.com

Grammaticality judgements (GJs) are frequently used by syntacticians to ascertain whether or not a particular construction is permitted in a language. This poster describes an experiment that tested whether GJs could be affected by phonetic detail. A socially-meaningful phonetic variable in New Zealand English (NZE) was manipulated in grammatically standard and non-standard constructions. There were significant differences between the ratings given to the sentence with a conservative realisation of the variable and an innovative realisation, and the size and direction of these differences was dependent on, amongst other factors, the ability of the variants to alter the class rating of speakers. Not only are the methodological implications of this study non-trivial, but the findings also pose questions about the storage and access of constructions.

Fifty sentences that ended with a phrase-final /t/ were specially designed and recorded for the experiment. Phrase-final /t/ is a socially conditioned variable in NZE (Docherty et al., 2006) that is mostly realized as an unreleased, glottalized variant, and is most likely to be released if the speaker is an older, professional female. The sentences used contained a range of constructions: *good* sentences which were considered to have standard constructions, *bad* sentences which contained constructions that no native speaker would be expected to use, and three socially variable constructions - preterite *come* and *done* and possessive HAVE-*got* - which have all been documented as being more frequent in the speech of younger, non-professionals (Quinn, 2000). The fifty recordings used in the experiment were ones where the five female speakers had produced glottalization before also releasing the final /t/. By manually cutting the release from the recording, we had a sentence with the innovative, unreleased variant, and the original uncut recording had the opposing conservative, released variant. Two pilot studies (one is described in Walker, 2007) that included both versions of the recordings showed that the age and social class ratings of the speakers were significantly higher with the conservative variant.

In this experiment, participants were asked to rate the grammaticality of the sentences on a six point scale, where 6 meant the sentence was 'completely grammatical' and 1 meant that it was 'completely ungrammatical'. The two versions of the manipulated sentences were presented dispersed between an additional sixty filler sentences. Thirty people participated in the experiment.

The data was analyzed with a hand-fit linear regression model. The age and social class ratings given to the sentences in one of the pilot studies were included as independent effects. Included in factors that affected the overall grammaticality ratings of the manipulated sentences was the average age rating the particular token had garnered in the pilot. This worked such that the older the rating of a speaker in a particular sentence, the more grammatical the sentence was rated with the innovative variant.

To explicitly test the size and direction of any differences in the rating of the conservative compared to the innovative version of a sentence, the difference between the two was put into another series of linear regression models as the dependent effect. A factor that was significant in the size and direction of the difference was how much the released /t/ had raised the class ratings in the pilot. The more a token had garnered a higher social class rating for the speaker with the conservative compared to the innovative realisation, the more people rated the sentence as more grammatical with the innovative variant. When we looked at each type of sentence individually, this effect seemed to be being carried primarily by the socially variable *come* sentences.

The results suggest that participants found the sentences more grammatical when they believed the speaker was younger or of a lower social class (class and age were highly correlated). For the same reason, in the same way, the socially salient realisation of the phrase-final /t/ affected the grammaticality ratings of the sentences. The fact that the effect appears to be carried primarily by the *come* sentences, which are said mostly by younger, non-profesional speakers, suggests the direction of the effect could be due to the way participants treat encountered constructions: they store them complete with speaker information.

The grammaticality of sentences is not usually considered in terms of speaker or phonetic factors. This study shows that socially-meaningful phonetic variation can alter the grammaticality ratings given to sentences in a predictable fashion. This poses serious methodological issues for those who would present stimuli for judgement in an auditory form (see Kitagawa & Fodor, 2006), because judgments could be affected by such extra-linguistic information. One could also argue, however, that the social and phonetic effects seen here are not in fact external to the grammar, but rather that this sort of information is intrinsically tied to how we store, access and process constructions.

References

Docherty, G., Hay, J., & Walker, A. (2006). Sociophonetic patterning of phrase-final /t/ in New Zealand English. In P. Warren & C. Watson (Eds*.), Proceedings of the 11th Australasian International Conference on Speech Science & Technology* (pp. 378-383). Canberra: Australian Speech Science & Technology Association Inc.

Kitagawa, Y., & J. D. Fodor (2006). Prosodic Influence on Syntactic Judgements. In G. Fanselow, C. Fery, R. Vogel & M. Schlesewsky (Eds.), *Gradience in Grammar* (pp. 336– 358). Oxford: Oxford University Press.

Quinn, H. (2000). Variation in New Zealand English syntax and morphology. In A. Bell & K. Kuiper (Eds.), *New Zealand English* (pp. 173-197). Wellington: Victoria University Press.

Walker, A. (2007). The effect of phonetic detail on perceived speaker age and social class. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 1453-1456). Saarbrucken, Germany.

Vowel Cooccurrence in the Lexicon: An Anti-OCP Effect?

Mary Ann Walter

Department of Linguistics, Northwestern University, USA; m-walter@northwestern.edu

While gradient cooccurrence avoidance of similar or identical consonants within lexical items is well-documented, similar investigations of vowel patterns are lacking. Phonetic patterns suggest that an opposite tendency may exist, and lead to observed asymmetries in the phonological patterning of vowels and consonants. Results from the lexicostatistics of Croatian and Spanish support this conjecture.

Background

Vowels behave qualitatively differently from consonants in a number of ways. One such difference is a marked asymmetry in categorical phonological cooccurrence patterns within lexical constituents. Vowels are much less likely to dissimilate from each other than consonants – constituting only 8 of 57 examples in Suzuki's (1998) cross-linguistic survey of dissimilation – and dissimilate in more local, restricted contexts than consonants when they do. Conversely, harmony processes enforcing assimilation/identity along some featural dimension apply overwhelmingly to vowels rather than consonants. These phonological asymmetries are paralleled by a line of research showing extensive coarticulation across vowels, regardless of intervening consonants (e.g. Öhman, 1966; Linebaugh, 2007). Taken together, these facts suggest that an articulatory tendency toward coarticulation among vowels may lead to asymmetries in vowel cooccurrence within a lexicon, eventually resulting in the widespread existence of phonological harmony (and relative lack of dissimilation). Surprisingly, however, patterns in vowel cooccurrence have hardly been investigated. This contrasts with consonant cooccurrence patterns, for which avoidance of identical and highly similar segments in close proximity has been documented in over forty languages, and claimed to be universal. The current study attempts to address this gap, and tests the prediction of a gradient tendency toward harmony in vowel cooccurrence patterns through corpus studies of Croatian and Spanish.

Method

The Croatian study draws on the Ukrstenko corpus of Bosnian/Croatian/Serbian wordforms (Sipka, 2008), which proceeds alphabetically through the letter 'h.' A lemmatized list of slightly over 20,000 unique root forms was generated from this corpus, excluding monovocalic ones, proper nouns, and most derivationally complex forms (n=21,031). The Spanish study is based on a list of unique word forms, including derived but not inflected forms, drawn from the Lexesp corpus (Sebastian, Cuetos, Marti & Carreiras, 2000). The resulting list is comprised of roughly 20,000 items (n=21,412).

The grep software tool was then used to identify forms including all possible vowel pairs, regardless of ordering, separated by either one or two wildcard characters, or anywhere within a word (Spanish searches were limited to type 1). The same computation was performed in Croatian for the voiceless stop consonants [ptk] in CVC sequences as well as for the vowels, and in Spanish for the two voiceless stops [pt]. Observed versus expected cooccurrence was then computed according to the method pioneered by Frisch, Pierrehumbert and Broe (2004), which controls for phoneme frequency. A value of less than one indicates underattestation of a particular phoneme pair – the numerator is smaller than the denominator, indicating that fewer tokens occur than would be expected based on random free permutation of segments given their individual frequencies. Values greater than one indicate overattestation. As a statistical test of the difference in cooccurrence rates, the Bernoulli process was carried out over the vowel and consonant data in each language. Each diphone in each word form constitutes a single Bernoulli trial, with success defined as diphone cooccurrence of identical segment types, and failure as a non-identical pair.

Results

Croatian demonstrates the familiar effect of cooccurrence avoidance for identical consonants (Bernoulli process $p<0.0001$). The drastic nature of this asymmetry is not reflected in the vowel data, however. Table 1 gives mean observed/expected ratios for vowel pairs according to the three string types listed above, as well as the consonant data. Values cluster much more tightly around the null hypothesis value of 1 (indicating free cooccurrence) than was seen for consonants both within Croatian, and cross-linguistically. For the VCV string type, identical vowel cooccurrence is still significantly underattested relative to what is expected ($p=0.01$). However, the degree of underattestation is drastically less for vowels than for consonants. Despite overall underattestation, identical vowels are much more likely to cooccur in Croatian than identical consonants are.

Table 1: Croatian and Spanish observed/expected ratios.

	CROATIAN				SPANISH	
	CVC	VCV	VCCV	V…V	CVC	VCV
Identical	.79	.98	.97	.98	.74	.82
Non-Identical	1.35	1.01	.99	1.02	1.06	1.05
Difference	.56	.03	.02	.04	.32	.23

Spanish displays a similar, though less marked, asymmetry. Once again, for the voiceless stop consonants cooccurrence is observed less often than expected, and the result is highly significant ($p<0.0001$). And again, the vowels show significant underattestation of identical diphones across a single segment, but to a much lesser degree than consonants ($p<0.0001$).

Discussion and Conclusions

While no statistical tendency toward vowel harmony is observed in these two languages, the tendency toward enforced *dis*harmony is considerably weaker for vowels than for consonants in both. A more detailed analysis with multiple similarity classes might exhibit a stronger effect. Alternatively, the (small) vowel inventory size of both languages may be playing a role, as it is known to affect perception of similarity (Hacquard, Walter & Marantz, 2007).

As predicted, vowels behave differently from consonants with respect to cooccurrence patterns. Their different patterning may stem from the different information they convey, which has been described as *indexical* rather than *lexical*. Due to their greater intensity and potential for compensation among articulators, vowels may be more important for speaker evaluation than lexical identification. This asymmetry eventually results in distributional properties within the lexicon that both reflect and potentially lead to categorical phonology.

References

Frisch, S., Pierrehumbert, J.B., & Broe, M. (2004). Similarity avoidance and the OCP. *Natural language and linguistic theory,* 22, 179-228.

Hacquard, V.C., Walter, M.A., & Marantz, A. (2007). The effects of inventory on vowel perception in French and Spanish: An MEG study. *Brain and language,* 100(3), 295-300.

Linebaugh, G. (2007). *Phonetic grounding and phonology: Vowel backness harmony and vowel height harmony*. Ph.D. dissertation, University of Illinois at Urbana-Champaign.

Öhman, S. (1966). Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America,* 39, 151-168.

Sebastián, N., F. Cuetos, M. A. Martí,, & M. F. Carreiras. (2000). *LEXESP: Léxico Informatizado del Español. Edición en CD-ROM.* Barcelona: Edicions de la Universitat de Barcelona (Colleccions Vàries, 14).

Sipka, D. (in progress). *Ukrstenko corpus of Bosnian/Croatian/Serbian.*

Suzuki, K. (1998). *A typological investigation of dissimilation*. Ph.D. dissertation, University of Arizona.

Do Chinese learners of English make use of fine phonetic details as English speakers do when perceiving the English lexical stress?

Qian Wang & Tae-Jin Yoon

Department of Linguistics, University of Victoria, Canada; wangqian@uvic.ca; tyoon@uvic.ca

Introduction

Lexical stress and phrasal accent comprise two levels of prominence in English. Phrasal accent is primarily cued by F0 while lexical stress is jointly cued by duration, intensity, and spectral properties (Sluijter & van Heuven 1996). Earlier studies with non-native speakers have shown that they are not very different from native speakers in their perception of lexical stress on words with phrasal accent, i.e. words in citation forms or focused/accented position in a sentence (cf. Davis & Kelly 1997, Adams & Munro 1978). Little is known, however, about how well they would discriminate lexical stress on words that bear no phrasal accent. Non-native speakers may differ from native speakers in utilizing aspects of phonetic detail in perceiving lexical stress. Thus, we compare the performance of Mandarin Chinese learners of English (CE) and native English speakers (NE) in the perception of words excised from ACCENTED and UNACCENTED conditions. Four possible outcomes are anticipated. First, if CE do not differ from NE in their perception of lexical stress on words taken from both conditions, then they arguably resemble NE in their use of the acoustic cues in lexical stress perception. Second, if CE show no difference from NE in the accented context but worse performance in the unaccented condition, then they may be seen as sensitive to F0 but less so to duration and intensity. Third, if CE show significantly worse performance than NE under the accented condition but show no difference under the unaccented condition, then it can be argued that CE are sensitive to duration and intensity cues but not F0. The last possibility is that CE are worse in both conditions, which would indicate that CE are insensitive to cues signaling both stress and accent.

Experiment Design and Results

To test these hypotheses, we designed a perception experiment using naturally produced sounds. We chose one word form with two possible stress patterns: PERmit as a noun and perMIT as a verb (where capital letters indicate the stressed syllable). Ten speakers produced each form of the word, which was embedded in two focus contexts: one where the target word is the focus of the sentence, and the other where a target word is not, i.e., where another content word in the sentence is under focus. For each combination of stress pattern and focus type, 4 different sentences were created. Each sentence was recorded twice, producing a total of 32 tokens (2 stress patterns × 2 focus conditions × 4 sentential types × 2 repetitions). Ninety-six words which were produced with the intended stress pattern and accent were excised from the sentences. These tokens were used to construct an oddity test, such that a participant listened to 3 tokens in a triad and decided whether the tokens are all the same and if not, which token was the odd one. Thirteen NE (7m, 6f) and 11 CE (6m, 5f) participated in the perception experiment. There were four different combinations of accented conditions, a triad with all three accented tokens (AAA), a triad of all unaccented words (UUU), a triad of two accented and one unaccented words (AAU) and a triad of one accented and two unaccented words (AUU). Here, A stands for accented and U for unaccented word. A total of 128 triads (32 triads of each accent combination) were constructed and divided into four blocks. A constraint in each triad was imposed upon the stimuli such that no two tokens were from the same speaker.

Error rates were collected from each participant and analyzed. A 2 (language group) × 4 (accent combination) mixed-model ANOVA revealed that the main effect for language group was significant ($F(1, 22) = 9.42$, $p < .01$, $\eta^2 = .30$). There was an overall difference in the error rate of Mandarin learners of English (M=21.4%) compared to native English speakers (M=10.3%). A significant main effect for the accent condition ($F(3, 22)=21.07$, $p<.001$, $\eta^2 = .48$), and a significant language group by accent interaction ($F(1,22)=9.11$, $p<0.01$, $\eta^2 = .29$) were also obtained. Two-tailed independent t-tests for each accent condition between native English speakers and Chinese learners indicated that there were significant differences in UUU and AAU, but not in AUU and AAA contexts (see Figure 1): UUU ($t(22)=3.71$, $p<0.005$; English: M=10.1%, SD=8%; Chinese: M=37.50%, SD=25.15%); AAU ($t(22)=2.13$, p=0.044<0.05; English: M=11.38%, SD=9.3%; Chinese: M=20.08%, SD=10.67%); AUU (t(22)=1.97, p=0.061>0.05; English: M=16.03%, SD=12.27%; Chinese: M=25.57%, SD=11.22%); and AAA (t(22)=.72, p>.4; English: M=3.85%, SD=0.06%; Chinese: M=2.27%, SD=0.04%).

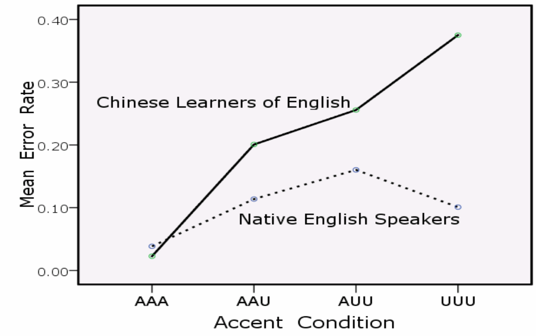


Figure 1: Interaction mean plot that shows error rates as function of accent effects by two language groups. Four types of accent combination are presented on the x-axis and error rates are represented on the y-axis. The performance of NE is indicated by dotted line, and that of CE is represented with straight line.

Discussion and Conclusion

Focusing on the AAA and UUU conditions alone, the results are consistent with our second hypothesis: that is, if CE show no difference from NE in an accented context but worse performance in an unaccented condition, then they may be seen as sensitive to F0 but less so to duration and intensity. CE's comparable performance with NE in perceiving lexical stress in accented contexts may be ascribed to their sensitivity to F0, whereas their difficulty with unaccented words may be due to their insensitivity to duration and intensity. Previous studies have shown that duration, intensity, and spectral properties are reliable cues for English stress perception by native speakers when F0 information is absent (Campbell 1993, Sluijter & van Heuven 1996). Second language (L2) learners of English may not have such a strategy for stress perception. Studies have shown that L2 learners have a perceptual bias toward the familiar cues in their first languages (Gandour 1983, Nguyen 2003). This asymmetrical pattern of performance is consistent with the first author's earlier experiment on the perception of lexical stress based on synthetic bisyllabic words with manipulated F0, intensity, and duration information (Wang 2008). Wang (2008) demonstrated that, in investigating L2 stress acquisition, CE are sensitive to F0 only, unlike NE speakers who are sensitive to intensity and duration as well as F0. It is not sufficient to ascertain whether there is stress deafness or not. Rather, we should investigate the fine phonetic detail involved in L2 stress perception. The present study raises questions on the perception of lexical stress by English learners with Japanese, Finnish, and Korean language backgrounds, which have different accentual or stress systems than Mandarin Chinese. It will be interesting to observe what phonetic details they would attend to in perceiving English stress.

References

Adams, C., & Munro, R.R. (1978). In search of the acoustic correlates of stress: Fundamental frequency, amplitude, and duration in the connected utterances of some native and non-native speakers of English. *Phonetica*, 35(3), 125-156.

Campbell, N. (1993). Automatic detection of prosodic boundaries in speech, *Speech Communication*, 13(3-4), 343-354.

Davis, S.M., & Kelly, M.H. (1997). Knowledge of the English noun-verb stress difference by native and nonnative speakers. *Journal of Memory and Language*, 36(3), 445-460.

Gandour, J.T. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics*, 11, 149-175.

Nguyen, T.A.T. (2003). *Prosodic Transfer: The Tonal Constraints on Vietnamese Acquisition of English Stress and Rhythm.* Doctoral dissertation, University of Queensland, Australia.

Sluijter, A.M.C., & van Heuven, V.J. (1996). Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America*, 100(4), 2471-2485.

Wang, Q. (2008). L2 stress perception: The reliance on different acoustic cues. In *Proceedings of the 4th International Conference on Speech Prosody*, Campinas, Brazil.

Ka conversion – the changing sound and rhythm of Māori?

Catherine Watson[1], Margaret Maclagan[2], Ray Harlow[3], Winifred Bauer[4], Jeanette King[2] and Peter Keegan[1]

[1] University of Auckland, New Zealand, [2] University of Canterbury, New Zealand, [3] University of Waikato, New Zealand, [4] Victoria University of Wellington, New Zealand
c.watson@auckland.ac.nz, margaret.maclagan@canterbury.ac.nz, rharlow@waikato.ac.nz, j.king@canterbury.ac.nz, winifred.bauer@vuw.ac.nz, p.keegan@auckland.ac.nz

Introduction

Māori is the language of the indigenous people of New Zealand. It has been in increasingly close contact with English for the last 180 years. The Māori and New Zealand English (MAONZE) project is tracing changes in the pronunciation of Māori over time. The phonemic long and short vowel pairs are becoming more similar (Maclagan et al., 2004, Watson et al., 2006) as the short vowels become more peripheral and the duration of the long vowels lessens. But the open vowel pair, WĀ /a:/ and WAKA /a/, retain their relative lengths even though the formant differences have lessened (Harlow et al., in press).

Biggs (1961:6), whose ground-breaking scholarship laid the foundation for all aspects of the modern linguistic analysis of Māori, lists ten particles which have short long variation. All except two occur phrase initially. The most frequent is the verbal particle *ka* (inceptive) (Bauer, 1993: 535; Biggs, 1961: 29), which is described as having two allomorphs, /ka:/ when the phrasal material following it contains no more than two morae, and /ka/ elsewhere (Biggs, 1973: 28). Observation indicated that younger speakers were not obeying the '*ka*-rule' and therefore losing this particular allomorphy. This study investigates the *ka*-rule over time, and compares it with already established vowel system changes.

Methodology

The research involves four groups of speakers: seven born in the 1880s (MU), five kaumātua (K, elders) born in the 1920s and 30s, six young speakers, three whose first language is Māori (L1Y) and three whose first language is English (L2Y), all born in the 1970s and 1980s. Up to 30 examples of potentially long and 30 examples of potentially short *ka* were analysed for each speaker, with potential length determined by the original rule. MAONZEminer (see http://www.ling.canterbury.ac.nz/onze/) was used to extract the tokens. PRAAT version 4.125 (Boersma and Weenink, http://www.fon.hum.uva.nl/praat/) was used to extract the first three formant values from the centre of the steady state portion of the vowel. The formant positions were visually checked and corrected as necessary. Length measurements were calculated from wide-band spectrograms together with the waveforms. The length of the vowel was determined from the start of voicing after the /k/ to the start of the following consonant. If *ka* was followed by a word starting with a vowel, length was taken to the middle of the vowel transition.

Results

Analysis confirms the operation of the '*ka*-rule' in the speech of the MU group (n= 7), for whom the length difference between /ka:/ (mean = 127.66, sd = 35.13) and /ka/ (mean = 62.17, sd = 24.35) is clearly distinct with an effect size (Cohen's D) of 2.20. The difference is less but significant for the K group (n= 5), where the figures are /ka:/ (mean = 94.70, sd = 30.07) and /ka/ (mean = 62.89, sd = 27.53), with an effect size of 1.03 ($t = 7.99$, $df = 242$, $p < 0.001$). Only one younger speaker (an L1 speaker) has a small but significant difference between /ka:/ (mean = 64.74, sd =22.83) and /ka/ (mean = 49.08, sd =19.02) with an effect size of 0.75 ($t = 2.20$, $df = 38$, $p = .034$). The other five younger speakers lack difference and therefore do not use the '*ka*-rule'. In contrast the duration of WĀ is significantly longer than WAKA for all speaker groups. All MU and most K speakers make /ka/ and WAKA significantly more central in F1-F2 space than /ka:/ and WĀ. However whilst all L1Y and L2Y separate WĀ and WAKA in length and F1-F2 space, they do not separate /ka:/ and /ka/ with both closer to their WAKA than their WĀ in length and F1-F2 space. Histograms show that no group has bimodal plots for /ka:/ or /ka/. The /ka:/ lengths shorten until they totally overlap the /ka/ distribution for the L2Y speakers.

Discussion

The results show that the allomorphy attested for in *ka* in the older groups no longer exists for most of the younger speakers, despite the fact that elsewhere, the distinction between WĀ/WAKA has, alone of the vowel pairs, resisted length neutralisation in that same period. We ask why the *ka*-rule has broken down, if WĀ/WAKA resists length neutralisation. We suggest that the loss results from the general weakening of the distinction between long and short vowels, combined with the non-contrastive allomorphic context, together with unstressing of common grammatical words (English influence, see Bybee, 2002). There has been a break in the transmission of Māori between the MU speakers and the young speakers. While there is now a growing number of young, fluent Māori speakers, it is not known how many are L1 Māori speakers. However it is clear that a significant number of the young speakers have learnt from people for whom Māori is a second language.

There are many fewer contexts for long *ka* than short *ka* in modern Māori (only 187 in the first 1,000 *ka* contexts in Boyce, 2006). The MU speakers produced similar numbers of long and short *ka* contexts, but the number of long *ka* contexts decreases from the K to the L2Y speakers. As long *ka* becomes less common, there are fewer instances to learn from. However our observations suggest that length is retained in a few common long contexts such as *ka pai* and *ka nui*. Anecdotal evidence suggests that the rhythm of Māori is being influenced by NZE just as NZE has increasingly influenced Māori vowels over time (Harlow et al., in press; Maclagan et al., 2004). This impact is likely to be significant since English is stressed-timed whereas Māori is conventionally regarded as mora-timed (Bauer, 1981). A shift from mora-timing as the basis of Māori prosody involves a shift to a syllable-based rhythm, since the syllable is the relevant prosodic unit for stress. That will inevitably weaken the *ka*-rule, because 2-morae and 3-morae words frequently have the same number of syllables (e.g. *haere* has 3 morae, but two syllables). In order to survive, the *ka*-rule would need to become lexical. In addition, the traditional *ka*-rule (Biggs, 1973: 28) is the reverse of the rules for other initial particles (such as the possessive prepositions *no, na, mo, ma*) which agree in length with the following syllable. Similarly, the basic principle that the minimal phrase in Māori consists of 3 morae (Bauer, 1997: 451) would not require alternation in the length of *ka* because all lexical bases contain at least two morae; rather the *ka*-rule produces phrases with a minimum of four morae. Because it is so exceptional, the *ka*-rule is more likely to be lost with the break in transmission.

Acknowledgments

This research was supported by a grant from the Marsden Fund of the Royal Society of New Zealand.

References

Bauer, W. (1981). Hae.re vs. ha.e.re: a note. *Te Reo*, 24, 31-36.

Bauer, W. (1993). *Māori*. London: Routledge.

Bauer, W. (1997). *The Reed Reference Grammar of Māori*. Auckland: Reed.

Biggs, B. (1961). The structure of New Zealand Maaori. *Anthropological Linguistics*, 3(3), 1-54.

Biggs, B. (1973). *Let's Learn Māori* 2nd edition. Auckland: Reed.

Boyce, M. (2006). *A Corpus of Modern Spoken Māori.* Unpublished PhD thesis, Victoria University of Wellington, NZ. Including *The Māori Broadcast Corpus on CD*.

Bybee, J. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change*, 14, 261-290.

Harlow, R., Keegan, K., King, J., Maclagan, M., & Watson, C. (in press). The changing sound of the Māori language. In J. Stanford & D. Preston (Eds.). *Quantitative Sociolinguistic Studies of Indigenous Minority Languages.* Amsterdam: John Benjamins.

Maclagan, M., Harlow, R., King, J., Keegan, P., & Watson, C. (2004). New Zealand English influence on Māori pronunciation over time. *Te Reo*, 47, 7-27.

Watson, C., Maclagan, M., King, J., & Harlow, R. (2006) Are there L1 and L2 effects in the speech of young speakers of Māori? In P. Warren & C. I. Watson (Eds.), *Eleventh Australasian International Conference on Speech Science and Technology* (pp. 317-322). Auckland: Australasian Speech Science and Technology Association Inc.

Heads you lose: prosodic structure and timing in Hungarian

Laurence White* & Katalin Mády **

* Department of Experimental Psychology, University of Bristol; laurence.white@bristol.ac.uk
** Institute of Phonetics and Speech Processing, University of Munich; mady@phonetik.uni-muenchen.de

Prosodic structure influences speech segment duration. In English, for example, both heads and edges of words and higher-level prosodic constituents may be lengthened: stressed syllables are longer than unstressed syllables (e.g. Dauer, 1983); accented syllables are longer still, with additional lengthening in other syllables of the accented word (e.g. Turk & White, 1999); word-initial consonants are longer than word-medial consonants (e.g. Oller, 1973); phrase-final syllable rhymes are lengthened (e.g. Wightman, Shattuck-Hufnagel, Ostendorf & Price, 1992). Final lengthening in higher-level prosodic constituents is widespread in other languages as well, and possibly universal, but domain-head effects may vary between languages (Beckman, 1992). For example, Castilian Spanish shows minimal stress-based lengthening compared to English, and little or no evidence of accentual lengthening (Ortega-Llebaria & Prieto, 2007).

Hungarian presents an interesting perspective from the point of view of speech timing. Traditionally regarded as a "syllable-timed" language, it has a phonological length contrast for vowels (Siptár & Törkenczy, 2000). This suggests a constraint on prosodic timing effects: short vowels may not be lengthened, or long vowels shortened, so much as to cause perceptual ambiguity at the segmental level. Indeed, it has been claimed that intensity variation, not vowel lengthening, is the primary cue to lexical stress in Hungarian (Fónagy, 1958).

Lexical stress is, in theory, a wholly reliable cue to word boundaries in Hungarian, as it is placed word-initially with complete consistency, and this may render other cues to lexical structure superfluous. Some studies have suggested that there is an influence of lexical structure on segment duration, however, an inverse relationship between word length and stressed vowel duration (Meyer & Gombocz, 1909; Tarnóczy, 1965). Targets in these studies were uttered as isolated words, however, so the observed polysyllabic shortening could be due to the attenuation of utterance-final lengthening in the longer words. Indeed, there is evidence from a single-speaker study comparing whole-word durations in utterance-initial, medial and final positions that indicates considerable final lengthening for monosyllables (Olaszy, 2006). The amount of lengthening was attenuated according to word length, but still significant for pentasyllables; however, data are lacking on the distribution of final lengthening within the word.

We investigated three potential influences of prosodic structure on the duration of phonologically long and short word-initial stressed vowels:

- Presence or absence of pitch accent.
- Utterance position: medial vs final.
- Word length: monosyllabic, disyllabic, trisyllabic.

We constructed sentences containing target words in which vowel length, accent, word length and utterance position were systematically varied. We recorded ten native speakers of Hungarian from Budapest reading these sentences three times, and analysed variation in the duration of long and short vowels according to the other prosodic factors.

As shown in Figure 1, we found a large effect of utterance position, with both long and short vowels lengthened by about 40% in absolute-final syllables. There was an influence of phonological vowel length on the distribution of final lengthening: only long vowels were lengthened – by 23% – in utterance-penultimate syllables.

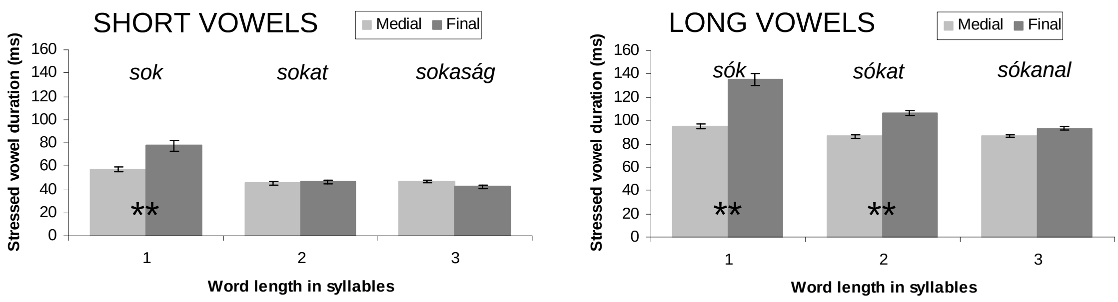


Figure 1: Mean vowel durations for short stressed vowels (left) and long stressed vowels (right) according to word length and utterance position. Error bars indicate ± one standard error. Significant differences between utterance-medial and utterance-final position are indicated ** ($p < 0.01$).

Lengthening of the stressed syllable in pitch-accented words appeared absent in Hungarian. Furthermore, as indicated by vowel durations in utterance-medial words (Figure 1), we did not find support for an inverse relationship between word length and vowel duration, suggesting that previous studies may have confounded the effects of word length and utterance position.

The overall picture in Hungarian prosodic timing thus provides some support for the idea that languages with a low durational marking of lexical stress tend not to manifest higher-level domain-head effects, in particular, accentual lengthening. Domain-edge effects, in contrast, may be universal, though the locus of final lengthening is here modulated by phonological vowel length. Further, the shortest utterance-medial long vowels (87 ms) were little longer than short vowels in absolute-final syllables (78 ms), suggesting that lengthening alone may be an insufficient perceptual cue to both phonological vowel length and utterance boundary location.

References

Beckman, M. E. (1992). Evidence for speech rhythms across languages. In Y. Tohkura, E. Vatikiotis-Bateson & Y. Sagisaka (Eds.), *Speech Perception, Production and Linguistic Structure* (pp. 457-463). Oxford: IOS Press.

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, 11, 51-62.

Fónagy, I. (1958). A hangsúlyról [On stress]. *Nyelvtudományi Értekezések*, 18.

Meyer, E.A., & Gombocz, Z. (1909). *Zur Phonetik der Ungarischen Sprache* [*On the Phonetics of Hungarian*]. Upsala: Berlings Buchdruckerei.

Olaszy, G. (2006). *Hangidőtartamok és Időszerkezeti Elemek a Magyar Beszédben* [*Sound Durations and Temporal Structure in Hungarian Speech*]. Budapest: Akadémiai Kiadó.

Oller, D.K. (1973). The effect of position in utterance on speech segment duration in English. *Journal of the Acoustical Society of America*, 54, 1235-1247.

Ortega-Llebaria, M., & Prieto, P. (2007). Disentangling stress from accent in Spanish: Production patterns of the stress contrast in deaccented syllables. In P. Prieto, J. Mascaró & M.-J.Solé (Eds.), *Segmental and Prosodic issues in Romance Phonology* (pp. 155-176). Amsterdam: John Benjamins.

Siptar, P., & Torkenczy, M. (2000). *The phonology of Hungarian*. Oxford: Oxford University Press.

Tarnóczy, T. (1965). Can the problem of automatic speech recognition be solved by analysis alone? *Rapport du 5. Congrès International d'Acoustique*, vol. 2, 371-387.

Turk, A.E., & White, L. (1999). Structural influences on accentual lengthening in English. *Journal of Phonetics*, 27, 171-206.

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America*, 91, 1707-1717.

That elusive rhythm: Pros and cons of rhythm metrics

Laurence White*, Sven L. Mattys*, Izabelle Grenon & Lukas Wiget***

***** Department of Experimental Psychology, University of Bristol, UK;
laurence.white@bris.ac.uk; sven.mattys@bris.ac.uk; l.wiget@bris.ac.uk
****** Department of Linguistics, University of Victoria, Canada; izabelle@uvic.ca

There has been a burgeoning of interest in empirical measures of speech rhythm in recent years. Following Dauer's (1983) elucidation of the differences in syllable structure between so-called "stress-timed" and "syllable-timed" languages, the main effort to quantify rhythm has focused on variation in vowel and consonant duration. Ramus, Nespor and Mehler (1999) utilised standard-deviation measures of interval variability (ΔV, ΔC) and a measure of durational balance of vowels and consonants within utterances (%V). Low, Grabe and Nolan (2000) proposed a Pairwise Variability Index (PVI), attempting to capture the sequential nature of rhythmic contrast.

Some of these metrics have been shown to correlate with speech rate, making them potentially unreliable (e.g. Barry, Andreeva, Russo, Dimitrova & Kostadinova, 2003). Dellwo and Wagner (2003) normalised for rate by dividing the standard deviation of interval duration by the mean: VarcoV for vowels, VarcoC for consonants. White & Mattys (2007) showed that VarcoV and %V were robust to speech rate variation, and were furthermore discriminative between languages held to differ rhythmically, and between first and second language speakers. Questions remain, however, about the reliability, scope and relevance of rhythm metrics:

- How consistent are rhythm metrics between different linguistic materials and measurers?
- How are rhythm metrics affected by durational processes unrelated to stress?
- Given that perception is a fundamental aspect of rhythm, how do rhythm metrics relate to listeners' experience of language?

Here, we present three studies designed to address these questions, with the intention of illuminating the scope and limitations of rhythm metrics, which are increasingly popular instruments in the field of laboratory phonology.

Study 1. Reliability of rhythm metrics

Rhythm metrics have usually been applied to read sentences, so questions remain about their applicability to spontaneous speech, which has many more hesitations and disfluencies, and may be produced at a different rate and with a different style from read speech. We compared rhythm scores for five scripted utterances with those for five phrases produced by the same speakers in an unscripted map description task. Scores for all metrics were strongly influenced by speaking style: importantly, differences in scores between spontaneous and read speech were of the same order of magnitude as the effects of language variety, indicating that comparisons between speaking styles are problematic. Inter-measurer reliability for read speech is high, however, although comparisons of absolute values between measurers must be considered with caution.

Study 2. Rhythm metrics and non-stress languages

Metrics have primarily been applied to the Romance and Germanic languages of Europe, language families which differ quantitatively in terms of stress production, particularly in the magnitude of stress-related vowel lengthening. Studies such as White & Mattys (2007) have demonstrated that rhythm metrics are informative about non-native adjustments between Romance and Germanic languages, showing vowel duration scores for second-language speakers to be intermediate between those of their native and target languages. Japanese lacks lexical stress altogether, however, but does have a durational contrast between phonologically long and short vowels. We examined the influence of first language on second language rhythm scores in the case of Japanese speakers adapting to Canadian English. We found that VarcoV did not differ for L1 and L2 speakers of English, suggesting good adaptation of the Japanese speakers to English stress patterns. Analysing ratios of specific stressed and unstressed vowels, however, we found significantly greater stress-related vowel lengthening by English L1 speakers than by Japanese speakers of English. This is precisely the sort of variation that rhythm metrics should capture, but the Japanese speakers' durational marking of the long/short vowel contrast leads to rhythm scores that spuriously suggest a native-like stress realisation. This result indicates that rhythm metrics may have limited applicability to quantity languages in which vowel and consonantal interval durations vary for phonemic contrast rather than as a consequence of stress production.

Study 3. Rhythm metrics and perception

Given the lack of evidence for isochronous intervals in speech (Lehiste, 1977), rhythm may be considered as much a perceptual phenomenon as an aspect of speech production. We looked for perceptual correlates of the differences in rhythm scores that we have found between varieties of English, including standard southern British English (SSBE), Welsh Valleys English and Orkney Island English, the latter two having VarcoV and %V scores intermediate between SSBE and "syllable-timed" Castilian Spanish. We used resynthesised monotone *sasasa* speech to focus on temporal characteristics of utterances (cf. Ramus, Dupoux & Mehler, 2003). By comparing resynthesised utterances from different varieties of the same language, we controlled for the possibility of distinctions being cued by cross-linguistic differences in stress distribution, a property not explicitly captured by rhythm metrics. To control for additional speech rate and timing cues, utterances were normalised for rate and trimmed to ten utterance-medial syllables. Using an AAX discrimination task to determine which pairs of resynthesised languages can be distinguished, we found that, as predicted by VarcoV and %V scores, Welsh Valley English was differentiated from both SSBE and Spanish, but not from Orkney Islands English.

Conclusion

We found rhythm scores for metrics such as VarcoV to be reliable between measurers, and predictive of listeners' discrimination ability between regional varieties of English. However, comparisonof rhythm scores between different styles of speech is shown to be problematic and the Japanese L2 data highlight potential limitations to the applicability of rhythm metrics.

Finally, patterns of rhythm scores between and within languages strongly challenge the old "stress-timed" vs "syllable-timed" typological distinction. A framework in which languages vary along a gradient scale of durational stress contrast may be more predictive of the degree to which languages also differ in the exploitation of duration to indicate other aspects of prosodic structure.

References

Barry, W. J., Andreeva, B., Russo, M., Dimitrova, S., & Kostadinova, T. (2003). Do rhythm measures tell us anything about language type? In M. J. Solé, D. Recasens & J. Romero (Eds.), *Proceedings of 15th International Congress of Phonetic Sciences* (pp. 2693-2696). Rundle Mall: Causal Productions.

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, 11, 51–62.

Dellwo, V., & Wagner, P. (2003). Relations between language rhythm and speech rate. In M. J. Solé, D. Recasens & J. Romero (Eds.), *Proceedings of 15th International Congress of Phonetic Sciences* (pp. 471-474). Rundle Mall: Causal Productions.

Lehiste, I. (1977). Isochrony reconsidered. *Journal of Phonetics*, 5, 253-263.

Low, E. L., Grabe, E., & Nolan, F. (2000). Quantitative characterisations of speech rhythm: 'Syllable-timing' in Singapore English. *Language and Speech*, 43, 377–401.

Ramus, F., Dupoux, E., & Mehler, J. (2003). The psychological reality of rhythm classes: Perceptual studies. In M. J. Solé, D. Recasens & J. Romero (Eds.), *Proceedings of 15th International Congress of Phonetic Sciences* (pp. 337-342). Rundle Mall: Causal Productions.

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, 73, 265–292.

White, L., & Mattys, S. L. 2007. Calibrating rhythm: First language and second language studies. *Journal of Phonetics*, 35, 501-522.

Replicator dynamics of speech perception and categorization

Colin Wilson[1], Michael Wolmetz[2], Paul Smolensky[2]

[1]Department of Linguistics, UCLA, USA; colin@humnet.ucla.edu
[2]Department of Cognitive Science, Johns Hopkins, USA; mikew@jhu.edu, smolensky@cogsci.jhu.edu

A central question in cognitive modeling is whether different levels of representation influence one another *interactively*, through bidirectional feedback or resonance (e.g., Grossberg, 1976, 2003; McClelland & Elman, 1986; Elman & McClelland, 1988; Massaro & Cohen, 1991; Magnuson et al., 2003, Rapp & Goldrick, 2000, 2004). In this paper, we demonstrate that interactivity is not needed to account for the detailed, time-dependent pattern of speech perception and categorization found in recent eye-tracking experiments (McMurray & Spivey 2000). Given minimal assumptions, a non-interactive mathematical model based on the replicator equation of evolutionary dynamics (Nowak 2006) correctly predicts both traditional behavioral findings of speech identification, and that categorization over time displays an *evolving sigmoid* pattern. This model can be instantiated as a novel type of connectionist network in which the activity of a unit is updated by multiplying its current value and the sum of the incoming excitation. Unlike many proposed models, the *replicator* network does not depend on the unverified claim that acoustic or auditory representations are affected by top-down feedback from phoneme, lexical, or other levels. This paper contributes to the formal understanding of the role and necessity of interactivity, the influence of both phonemic and acoustic information, and is (as far as we are aware) the first to apply the parsimonious formal methods of evolutionary dynamics to the domain of on-line speech perception.

McMurray & Spivey (2000) presented participants with a 9-step /pa/-to-/ba/ continuum in which Voice Onset Time (VOT) ranged from -50ms to +60ms in equal-sized steps of approximately 12ms. Participants heard the stimuli while viewing a computer display with the two response possibilities (/pa/ and /ba/) in predictable locations. Mouse-choice categorizations and eye-gaze movements and locations were recorded from stimulus offset. The distribution of eye gazes at the latest time period is essentially indistinguishable from the distribution of mouse choices; both show the sigmoid-shape curve that is characteristic of speech categorization experiments, with a sharp category boundary at approx. +10ms.

The eye-gaze data at earlier time periods shows a gradual evolution from a relatively 'flat' profile, which departs only slightly from chance looking, through a series of increasingly 'sharp' sigmoids that culminates in the final categorization distribution (see Fig. 1, reproduced from McMurray & Spivey, 2000: Figure 7). Extensive curve-fitting by McMurray and Spivey suggests that, at every measured point in processing, the eye-gaze responses have a sigmoid shape with pivot point at the category boundary. They call the temporal evolution from flat to sharp gaze distributions, observed for the first time in their experiment, the *evolving sigmoid*. Our lab is in the process of collecting relevant patient data, as well as extending McMurray & Spivey's important finding to a different (vowel quality) continuum.

McMurray & Spivey (2000) propose an interactive connectionist model, the Hebbian Normalized Recurrence Network, that succeeds in capturing the evolving sigmoid. Similar in its mechanics to an iteratively (re)normalized version of McClelland & Elman's TRACE or Grossberg's ART model, this network has two qualitative properties that McMurray & Spivey identify as crucial to its success. First, the network is sensitive to the *statistical structure* of the two categories /p/ and /b/ in virtue of competitive Hebbian learning; we approximate the statistical structure as two widely-separated normal distributions that give the probability densities of VOT values in each category. Second, the network displays *competitive processing* that plays out over time; the probability of responding with one category or the other changes during the competition in a way that mimics the experimentally observed sigmoid evolution. McMurray & Spivey's modeling results appear to support interactivity (feedback from the phoneme level of representation to the acoustic/auditory level) in speech perception and categorization.

However, the *replicator equation* (Nowak, 2004) is a straightforward embodiment of time-dependent competitive processing that does not require interactivity/feedback between levels of representation. The equation is $\dot{x}_i = x_i \cdot (f_i - \phi)$, where $\phi = \Sigma_j (x_j \cdot f_j)$. In biological applications, x_i denotes the proportion of the population that is of type or species i, $\dot{x}_i$ is the time derivative of x_i, f_i is the fitness of species i, and ϕ is the average fitness of the population. Competition follows from the fact that populations are represented by probability distributions: because the total probability mass of the population is fixed at 1, growth (positive time derivative) of one species implies decrease of another.

In our application of the replicator equation, there is no population of competing *species*, but rather competing representations. We identify x_i with the level of *activity* of cognitive representation i and f_i with the *fit* between i and the auditory representation of the incoming stimulus. To model McMurray & Spivey's results specifically, we assume two categories /b/ and /p/, each of which is associated with its own normal distribution over the VOT range. The fit between category /x/ and the auditory representation of a stimulus with VOT y is defined as the probability of a narrow VOT range centered at y given the distribution of /x/. Once the fitness values of /b/ and /p/ are determined for a given stimulus, the replicator equation deterministically governs the competition between the categories over time. No further communication between the auditory and category representations, and in particular no interactivity or feedback, occurs. Figure 2 illustrates how the replicator model yields the evolving sigmoid pattern found in McMurray & Spivey's experiments. (For the purposes of this simulation, the mean VOTs and standard deviations for /b/ and /p/ were estimated from values reported in the phonetics literature.) This result supports the claim that sensitivity to statistical structure and competitive processing – but crucially not interactivity — are the key qualitative properties of speech perception and categorization.

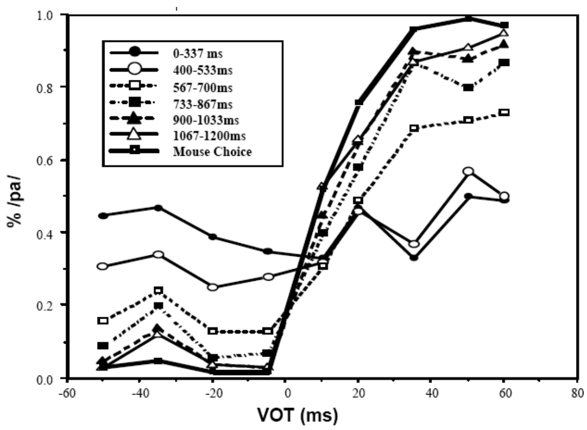


Figure 1: McMurray & Spivey (2000) results

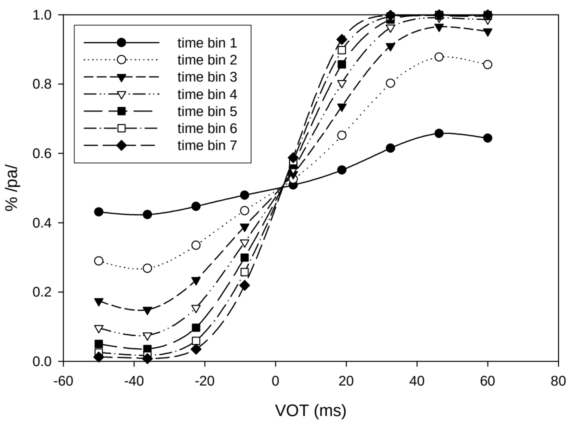


Figure 2: Predictions of the replicator model

References

Grossberg, S. (2003). Resonant neural dynamics of speech perception. *Journal of Phonetics*, 31, 423-445.

Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., & Dahan, D. (2003). The microstructure of spoken word recognition: Studies with artificial lexicons. *Journal of Experimental Psychology: General*, 132, 202-227.

Massaro, D. W., & Cohn, M. (1991). Integration vs. interactive activation: The joint influence of stimulus and context in perception. *Cognitive Psychology*, 23, 558-614.

McClelland, J L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18, 1-86.

McMurray, B., & Spivey, M. (2000). The categorical perception of consonants: The interaction of learning and processing. *Proceedings of the Chicago Linguistics Society*, 34(2), 205–220.

Nowak, M. A. (2006). *Evolutionary Dynamics: Exploring the Equations of Life*. Cambridge, MA: Belknap Press.

Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review*, 107, 460-499.

Can auditory distractors disrupt speech execution?

Ivan Yuen*, Kathy Rastle & Marc Brysbaert

Department of Psychology, Royal Holloway, University of London;*ivan.yuen@rhul.ac.uk

Evidence regarding the relationship between phonological encoding and motor execution in speech production is contradictory. Some studies suggest that these processes are staged with respect to one another (e.g., Damian 2003), while others suggest that they are cascaded (e.g., Goldrick & Blumstein, 2006; Kello, Plaut & MacWhinney 2000). For example, Goldrick and Blumstein (2006) argued for a cascaded model of speech production on the basis of tongue-twister data. They found that /k/ phonemes incorrectly produced as /g/ phonemes had longer VOTs than /g/ phonemes that were correctly produced. This result suggests that errors show phonetic traces of their intended (but unselected) targets. Similarly, Kello et al. (2000) reported that participants under time pressure exhibited longer response durations in a Stroop interference task when the targets and distractors were incongruent, once again supporting a cascaded model. However, Damian (2003) failed to replicate this result in his studies of picture-word interference and Stroop tasks, and thus argued for a staged model of speech production.

Our study contributes to this debate by examining the effect of unselected auditory distractors on the production of visually-presented nonword targets. Cascaded models predict that activation from the auditory distractors will 'leak' down into the motor system and thus distort the articulation of the nonword targets. Staged models predict no effect of the auditory distractors on the articulation of nonword targets.

Two tempo-naming experiments were conducted. In both experiments, subjects heard a sequence of five 50-millisecond beeps at regular intervals of 500 milliseconds. Participants were presented with a visual target on the fourth beep and were instructed to produce it on the fifth beep. Congruent or incongruent auditory distractors were presented immediately before (Experiment 1) or after (Experiment 2) the target. Congruent distractors had identical phonemes to the visually-presented targets, while incongruent distractors differed from the visually-presented targets by one phoneme. Participants were randomly assigned to two participant groups with different targets for production in each experiment. Target and distractor stimuli were classified into five types: (1) voicing of word-initial plosives (e.g., toop – doop); (2) voicing of word-final plosives (e.g., skite – skide); (3) additional phoneme /l/ to form a word-initial consonant cluster (e.g., pook – plook); (4) additional phoneme /w/ to form a word-initial consonant cluster (e.g., kig – kwig); (5) /l/ or /r/ in word-initial position (e.g., larg – rarg).

Our analyses examined various acoustic properties of targets (aspiration duration in (1), closure duration in (2), F1 and F2 in (3), F2 in (4) and F3 in (5)) in each of these conditions across the distractor congruency manipulation. Results from both experiments showed that productions from the two participant groups differed from one another as we expected. However, no significant differences were found between congruent and incongruent conditions within each participant group in terms of the measured acoustic properties. The congruency of auditory distractors had no influence on the production of target nonwords. We also measured the CV durations of the stimuli, and found no significant differences between the congruent and incongruent conditions. These findings challenge notions of cascaded articulation because they suggest that activation from unselected auditory distractors does not leak down to the motor system for speech execution.

Acknowledgements

This research was supported by a BBSRC grant (BB/E003419/1) awarded to Kathy Rastle and Marc Brysbaert.

References

Damian, M. (2003). Articulatory duration in single-word speech production. *Journal of Experimental Psychology,* 29, 416-431.

Goldrick, M., & Blumstein, S. (2006). Cascading activation from phonological planning to articulatory processes: Evidence from tongue twisters. *Language and Cognitive Processes,* 21, 649-683.

Kello, C., Plaut, D., & MacWhinney, B. (2000). The task dependence of staged versus cascaded processing: an empirical and computational study of Stroop interference in speech production. *Journal of Experimental Psychology,* 129, 340-360.

Harmful Reduction?

Frank Zimmerer*, Henning Reetz* & Aditi Lahiri#

*Institute for Phonetics, University of Frankfurt, Germany; zimmerer/reetz@em.uni-frankfurt.de
#Department of Linguistics, University of Oxford, England; aditi.lahiri@ling-phil.ox.ac.uk

In everyday conversation, talkers can be rather sloppy in their speech. Often, some segments are changed, some others are reduced, and yet others are deleted completely. Compared to “perfect” canonical speech or a citation form, natural speech is characterized by an enormous amount of variation. This is not to say that words are never produced close to a canonical pronunciation. But “massive” reductions occur commonly in natural speech (see, e.g., Johnson, 2004). The following German examples from the *Kiel Corpus of Conversational Speech* (IPDS, 1994) illustrate this point: *einverstanden* ‘agree$_{\text{PAST PARTICIPLE}}$’, *vielleicht* ‘perhaps’, or *irgendwie* ‘somehow’ were produced respectively as [nfɐʃtan̰], [ˈflaɪç] and [ˈɪ̰ɐŋv], with several segments deleted when compared to the citation forms [ˈʔaɪnfɐˌʃtandən], [fɪˈlaɪçt] and [ˈʔɪɐgəntˌviː]. Frequently, one encounters the statement that reductions and deletions are common in natural speech and that human listeners do not have difficulties in understanding what has been said. Given this logic, words like *einverstanden, vielleicht* or *irgendwie* produced as shown above, should be easily understood by native listeners of that language and hence, to enable them to do so, storing all variation in the mental lexicon is the only viable solution. In contrast, for models assuming a mental lexicon with only one very abstract representation for each word, such reductions would never be understood.

In the perception literature, naturally reduced items have been used in transcription tasks (e.g. Ernestus et al., 2002) and lexical decision experiments (e.g. Ernestus & Baayen, 2007). For priming studies examining natural variation and its impact on speech perception, experimental items were usually produced by trained speakers (e.g. McLennan et al., 2003, 2005; Sumner & Samuel, 2005; Tucker & Warner, 2007). We wanted to examine in more detail to what degree massively reduced words in natural speech affect comprehension. To our knowledge, a combination of different methods using transcription and lexical decision with priming of reduced words taken from natural speech has not been done before. We planned three experiments: (i) a transcription task of naturally spoken reduced and unreduced items, (ii) a cross-modal identity repetition priming experiment with lexical decision using items from Expt 1, and (iii) a transcription task of reduced and unreduced items in a sentential context followed by lexical decision priming with the same subjects.

In Expt 1, 92 word pairs, reduced and non-reduced forms of the same word, were split into two experimental lists. All utterances were taken from the Kiel Corpus which contains more than 2 hours of dialogues from 42 different speakers, with transcriptions by trained phoneticians (IPDS, 1994). The reduced forms were naturally produced, and the reductions and deletions were not based on any “regular” predictable phonological rules. The reduced items exhibited reductions of 1 or more segments. Their mean length was 6.38 segments (ranging from 3 to 13), consisting of 1 to 4 syllables (mean: 2.34). 16 (non-reduced) filler items were added for control, which were the same on both lists. 22 subjects transcribed what they heard, indicating how confident they were about their transcription. Results show that listeners had severe problems in understanding reduced words out of context. The reduced items were transcribed significantly worse (overall, 64% correct) than the non-reduced counterparts (overall, 94% correct). This was also reflected in the confidence ratings (6.8 for reduced vs. 9.2 for unreduced items).

From the first transcription experiment, 30 word pairs, where the reduced word was transcribed rather poorly, were used as auditory primes for visual targets in Expt 2. The reduced words had reductions of two segments or more. 30 control words that were not related to the targets were added. The experimental design was as depicted in (1). Subjects had to give a lexical decision (word/nonword) on the visually presented targets. Crucially, while unreduced auditory primes led to significant priming of their visual targets by 74 ms, reduced primes failed to produce target facilitation, (9 ms (n.s.) slower than the control items). Thus, reduced variants patterned like the unrelated controls and could not facilitate the recognition of the target word. These results suggest that reductions, familiar in natural speech, are not stored since they were not different from the unrelated control condition.

(1)	*Gloss*	*Prime (auditory)*		*Target (visual)*
	‘doctor’	$\text{Doktor}_{\text{REDUCED}}$	⟶	
	‘broadcast’	$\text{Sendung}_{\text{CONTROL}}$	⟶	DOKTOR
	‘doctor’	$\text{Doktor}_{\text{UNREDUCED}}$	⟶	

It has been shown that context helps to disambiguate reduced variants (e.g. Picket & Pollack, 1963; Pollack & Picket, 1963; Ernestus et al., 2002). Therefore, in the first part of Expt 3, we presented the items in their sentential context for transcription. We found that context indeed helped subjects to correctly transcribe the words. Another caveat of Expt 2 was that subjects had not heard the identical – albeit natural – pronunciation of each of the reduced (and unreduced) words, hence, it could be that there was no exact entry for the reduced items, and consequentially, no priming. Accordingly, we were also interested whether prior exposure to the reduced items affected recognition of the target. Thus, the second part was a priming study, identical set-up to Expt 2 (items without context). Again, we find that in spite of earlier exposure and the possibility to store a reduced exemplar, there was no priming. That is, although listeners were exposed to the reduced variants in context, the very same forms failed to prime later the target words.

These results lend support to the assumption that listeners do have an abstract word stored which can only tolerate predictable variations, but not massive, unpredictable reductions. An account positing different exemplars with fine phonetic detail (i.e. natural reduction) on a word base is not able to explain the results reported here. One possibility that has not been examined is that exemplars are not only stored with fine phonetic detail of the word, but also with their context. Further studies are needed to test whether such models are more successful in explaining the data that has been reported here. Words without context are to some extend unnatural. This is also true for the relatively unreduced items, where we found priming. However, so far, a model with abstract entries seems a more promising assumption to posit. It is also essential to examine deletions and their consequences for recognition cross-linguistically focusing on typological differences and their impact on these processes.

Acknowledgements

This research was supported by the German Research Council *DFG* (SPP 1234; SFB 471).

References

Ernestus, M., & Baayen, H. (2007). The comprehension of acoustically reduced morphologically complex words: the roles of deletion, duration, and frequency of occurrence. In Jürgen Trouvain & William J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 773-776). Saarbrücken, Germany.

Ernestus, M., Baayen, H., & Schreuder, R. (2002). The recognition of reduced word forms. *Brain and Language*, 81, 162-173.

IPDS – Institut für Phonetik und digitale Sprachverarbeitung (1994). *The Kiel Corpus of Spontaneous Speech, CD-ROM I-III*. Kiel: IPDS.

Johnson, K. (2004). Massive reduction in conversational American English. In K. Yoneyama & K. Maekawa (Eds.), *Spontaneous speech: Data and analysis* (pp. 29–54). Tokyo: The National Institute for Japanese Language.

McLennan, C. T., Luce, Paul A., & Charles-Luce, J., (2003). Representation of lexical form. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 29, 539-553.

McLennan, C. T., Luce, Paul A., & Charles-Luce, J., (2005). Representation of lexical form: Evidence from studies of sublexical ambiguity, *Journal of Experimental Psychology: Human Perception and Performance*, 31, 1308-1314.

Picket, J. M., & Pollack, I. (1963). Intelligibility of excerpts from fluent speech: Effects of rate of utterance and duration of context. *Language and Speech*, 6.151-164.

Pollack, I., & Pickett, J. M. (1963). The intelligibility of excerpts from conversation. *Language and Speech*, 6, 165-171.

Sumner, M., & Samuel, A. G. (2005). Perception and representation of regular variation: The case of final /t/. *Journal of Memory and Language*, 52, 322-338.

Tucker, B. V., & N. Warner (2007). Inhibition of processing due to reduction of the American English flap. In Jürgen Trouvain & William J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 1949-1952). Saarbrücken, Germany.

Laboratory Phonology 11

Victoria University of Wellington

New Zealand

30 June - 2 July 2008

Theme:

Phonetic detail in the lexicon

Sub-themes:

Accessing the lexicon
Invited speaker: **Keith Johnson, Berkeley**
Commentator: **Marcus Taft, UNSW**

Social information in the lexicon
Invited speaker: **Paul Foulkes, York**
Commentator: **Ben Munson, Minnesota**

Phonetic cues to lexical structure
Invited speaker: **Alice Turk, Edinburgh**
Commentator: **Laurie Bauer, VUW**

Generalising over the lexicon
Invited speaker: **Anne Cutler, MPI**
Commentator: **Karen Croot, Sydney**

LabPhon 11 LabPhoNZ

Victoria University of Wellington
Te Whare Wānanga o te Ūpoko o te Ika a Māui

UC University of Canterbury
Te Whare Wānanga o Waitaha
Christchurch New Zealand

ASSTA
The Australasian Speech Science & Technology Association Incorporated

REGISTRATION INFORMATION
&
ABSTRACT SUBMISSION DETAILS

are available from the conference website

http://www.victoria.ac.nz/labphon11/

Abstracts due 7th Dec 2007

For general enquiries please e-mail
labphon11@vuw.ac.nz

This conference is immediately followed at the same venue by the fifth conference of the International Gender and Language Association
3 - 5 July 2008 http://www.victoria.ac.nz/igala5/

Laboratory Phonology 11

Victoria University of Wellington

New Zealand

30 June - 2 July 2008

Theme:

Phonetic detail in the lexicon

Sub-themes:

Accessing the lexicon
Invited speaker: **Keith Johnson, Berkeley**
Commentator: **Marcus Taft, UNSW**

Social information in the lexicon
Invited speaker: **Paul Foulkes, York**
Commentator: **Ben Munson, Minnesota**

Phonetic cues to lexical structure
Invited speaker: **Alice Turk, Edinburgh**
Commentator: **Laurie Bauer, VUW**

Generalising over the lexicon
Invited speaker: **Anne Cutler, MPI**
Commentator: **Karen Croot, Sydney**

REGISTRATION INFORMATION
&
ABSTRACT SUBMISSION DETAILS

are available from the conference website

http://www.victoria.ac.nz/labphon11/

Abstracts due 7th Dec 2007

For general enquiries please e-mail
labphon11@vuw.ac.nz

This conference is immediately followed at the same venue by the fifth conference of the International Gender and Language Association
3 - 5 July 2008 http://www.victoria.ac.nz/igala5/