SECOND CONFERENCE ON LABORATORY PHONOLOGY

Outline Programme: 1.6.89

Part I: The Segment?

Day 1 - 30 June 89 (Friday)

Morning Papers: John Ohala (Berkeley): "The Segment: Primitive or Derived?"
                         Lieselotte Schiefer (Munich): "Trading Relations in the Perception of Stops
                         and their Implications for a Phonological Theory"

Discussants: Nick Clements (Cornell)
                         Björn Lindblom (Texas)
                         John Local (York)
                         Lisa Selkirk (UMass)

Afternoon Papers: Aditi Lahiri (Max Planck) and William Marslen-Wilson (APU Cambridge):
                         "Lexical Processing and Phonological Neutralisation"
                         Francis Nolan (Cambridge): "The Descriptive Role of Segments: Evidence
                         from Assimilation"
                         Jean-Marie Hombert and Gilbert Puech (Lyon II): Nasal Consonants and the
                         Development of Vowel Nasalisation

Discussants: Joan Mascaró (Barcelona)
                         John Ohala (Berkeley)
                         Cathe Browman (Haskins)

Day 2 - 1 July 89 (Saturday)

Morning Papers: Cathe Browman (Haskins) and Louis Goldstein (Yale):
                         "'Targetless' Schwa: an Articulatory Analysis"
                         Nigel Hewlett (QMC Edinburgh) and Linda Shockey (Reading): "Coarticulation
                         Increases with Rate?"

Discussants: Sarah Hawkins (Cambridge)
                         John Kingston (Cornell)

General Discussion of Part I with comments by Anne Cutler and Stephen Anderson.

Afternoon: Free
                         "Business meeting" with past and future organisers
                         Parasession on Declarative Phonology

Part II: Prosodic Structure

Day 3 - 2 July 89 (Sunday)

Morning Papers: Lisa Selkirk and Koichi Tateishi (UMass): "Syntax, Phrasing, and
Prominence in the intonation of Japanese"
Rob van den Berg, Carlos Gussenhoven, Toni Rietveld (Nijmegen): "Downstep
in Dutch: Implications for a Model"
Haruo Kubozono (Nanzan): "On the Hierarchy of Intonational Representation:
Evidence from Japanese"

Discussants: Mary Beckman (Ohio State)
Janet Pierrehumbert (Bell)
Jørgen Rischel (Copenhagen)
Bill Poser (Stanford)

Afternoon Papers: Janet Pierrehumbert and David Talkin (ATT Bell Labs): "Source
Dynamics as a Function of Prosodic Pattern"
Mary Beckman, Jan Edwards, Janet Fletcher (Ohio State): "Prosodic
Structure and Tempo in a Sonority Model of Articulatory Dynamics"

Discussants: Gunnar Fant (KTH Stockholm)
Louis Goldstein (Yale)
Osamu Fujimura (Ohio State)
Marina Nespor (Amsterdam)

General Discussion of Part II with comments by Irene Vogel and Gosta Bruce.

Part III:Phonology and Speech Technology

Day 4 - 3 July 89 (Monday)

Morning Papers: John Local (York): "Modelling Assimilation in Non-Segmental,
Rule-Free Synthesis"
Brian Pickering (IBM UK) and John Kelly (IBM UK and York): "Tracking
Long-Term resonance Effects"
Lou Boves (Nijmegen) Title to be announced

Discussants: Mario Rossi (Aix)
Klaus Kohler (Kiel)
Steve Isard (CSTR Edinburgh)

Afternoon: Possible Demonstrations at CSTR and/or Phonetics Lab

Posters

Posters describing recent work by Amalia Arvaniti, Anne Cutler, Gerry Docherty, Janet Fletcher, Sarah Hawkins, Jill House, Daniel Recasens, Jacques Terken, and Ian Watson will be on display in the open area at the back of the conference room from Friday till Sunday.

THE SEGMENT: PRIMITIVE OR DERIVED?

John J. Ohala

The essence of a segment is the close temporal coordination of
different articulators ("features") in order to create a distinctive
acoustic event.   Traditionally phonologists have regarded the segment
as a primitive.   Even when segments were broken down into
component features, the temporal coordination of these features
was still regarded as a given.   Autosegmental phonology now proposes
a kind of mixed view of the segment.   On the one hand segmentality
is recognized as a primitive in the CV (or X) tier but the features have
been dissociated from segments and exist on distinct atemporal tiers.
Considerable theoretical machinery has to be posited to get the
features linked up with the appropriate segments in the later stages
of a phonological derivation.
            In this paper I will argue that the primitive or nonprimitive
status of the segment (as defined above) needs rethinking.
I will argue that from an evolutionary point of view the temporal
coordination of different articulators does not have to be taken as
primitive; it is the probable consequence of the physical constraints of
the speech apparatus and the communicative task served by
speech.   Anatomical aerodynamic, and acoustic constraints dictate
how different articulators have to work together.   However, at the
present stage of speech, that is, in the way it is represented in
speakers' mental lexicons, this temporal coordination is best treated as
a primitive, that is, that most features are already linked together.
(The only exceptions might be the traditional suprasegmentals:
tone, intonation, etc.)   If so, then how can one account for features
like nasalization, pharyngealization, vowel palatality and the like
migrating through words and giving rise to such phonological phenomena
as nasal prosodies?   The answer is that these are historical relics of
the fact that though the temporal coordination of articulators is
required only in a relatively narrow time window, constraints of the
speech production mechanisms give rise to some temporal "spillover"
of articulator configuration outside that window.   Laboratory studies
show that listeners normally factor out this spillover and treat it
as a predictable distortion of the intended pronunciation.
Occasionally, however, listeners take the distortion at face value
and incorporate it as an intended part of their own pronunciation.
If other speakers copy this pronunciation this constitutes a sound
change.   However such a historical development is irrelevant to
today's speakers and thus is not part of the grammar of a language.

In sum, I maintain that the anchoring or spreading of features can
both be accounted for by the effect of phonetic constraints which operated
in history (whether of remote or close antiquity); at present both
behaviors should be treated as givens and not derived from more
primitive elements.   Similarly, according to Darwin, the sneer (lip
corner raised to expose the canines) originally served as a threat
signal because our distant ancestors had canines which were
disproportionately long and would thus serve to frighten others if
exposed to view.   However, to account for the signalling value of
a sneer today we do not have to posit in the sneerer's mental
conception of this gesture some "underlying" long canine teeth
(our canines are now no more awesome than any of our other teeth)
but the gesture of raising the lip corner by itself conveys mild
threat due to its constant association with other threat signals.

TRADING RELATIONS IN THE PERCEPTION OF STOPS AND THEIR IMPLICATIONS FOR A PHONOLOGICAL THEORY

Lieselotte Schiefer

Institut für Phonetik und Sprachliche Kommunikation der Ludwig-Maximilians-Universität München, F.R.G.

Hindi, an Indo-Aryan language of the Indian sub-continent, makes, unlike most other languages of the world, linguistic use of four stop classes: voiceless unaspirated, voiceless aspirated, voiced, and breathy voiced. LISKER/ABRAMSON (1964) failed to apply the concept of voice-onset time to the stops of Hindi, as only three classes (voiceless unaspirated, voiceless aspirated, and voiced) could unambiguously be defined on the VOT continuum, whereas the breathy stops seemed to share the same acoustical region (prevoicing or voicing lead) with the voiced ones. They claimed that the difference between the voiced and breathy voiced stops may be cued by 'other' acoustic features located in that acoustic portion following the release burst. On the other hand, LISKER/ABRAMSON noted that the voicing lead was absent in one of their speech samples. A missing voicing lead was reported, too, for Nepali (like Hindi an Indo-Aryan language) by POON/MATEER (1985), who found great interindividual differences in the realizations of breathy stops.

Phonologists have paid great attention to the question of how to represent the Hindi stops in terms of phonological features. Since CHOMSKY/HALLE (1968) several feature sets have been developed in order to account for the complexity of the Hindi stop system (e.g. HALLE/STEVENS 1971, LADEFOGED 1971, OHALA 1979, SCHIEFER 1984). In all feature sets only physiologically based features are used. None of the authors rely on results from perception tests.

This paper approaches the question of an adaquate featural representation of Hindi stops from an acoustic and perceptual point of view. The experiments, which will be discussed in detail, examined the problem of (i) whether breathy voiced stops are regularly produced with prevoicing by four Hindi informants and (ii) whether breathy voiced stops need prevoicing to be perceived as 'breathy' or whether they are unambiguously identifiable through acoustic features located in the release-burst and that portion following the burst. It turned out that there is no straight answer to this question since trading relations exist between both acoustic portions. We will discuss these results with respect to their implications for a phonological description of the stop system and whether Hindi stops are more adaquate represented by 'gestural'or 'articulatory phonology' as proposed by BROW-MAN/GOLDSTEIN (1986).

Lexical processing and phonological representation

Aditi Lahiri and William Marslen-Wilson

We are concerned with the mental representation of lexical items and the way in which the acoustic signal is mapped onto these representations during the process of recognition. We propose here a psycholinguistic model of these processes, integrating a theory of processing with a theory of representation. The cohort model of spoken word-recognition (Marslen-Wilson, 1984; 1987) is taken as the basis for our assumptions about the processing environment for lexical access. We assume further that the properties that phonological theory assigns to underlying representations of lexical form are functionally isomorphic to the listener's mental representations of lexical form in the "recognition lexicon", and that these representations have direct consequences for the way in which the listener interprets the incoming acoustic-phonetic information, as the speech signal is mapped onto the lexicon.

These claims were tested in two empirical investigations. The first involved the representation and spreading of a melodic feature (the feature [nasal]), and the second concerned the representation of quantity, specifically geminate consonants. In the first case, a cross-linguistic gating study was conducted to investigate how listeners interpret oral and nasal vowels in languages (English and Bengali) which contrasted the underlying status of the feature [nasal]. Listeners' performance, for both languages, was best explained by the hypothesis that representations of featural information in the recognition lexicon are abstract and underspecified. In the second study, we investigated the interpretation of quantity, asking whether closure duration -- the predominant perceptual cue to consonant gemination in Bengali -- was interpreted in the same way as melodic features (like nasality), or, as our theory predicted, in terms of information about the skeletal tier. The results showed that listeners do not reliably interpret duration as a cue to gemination until the appropriate stuctural information becomes available. We conclude that lexical processing is best understood in terms of very abstract perceptual representations, rather than representations which simply reflect the surface forms of words in the language. This has potentially wide-reaching consequences for the conduct of both psychological and phonological research into the representation of lexical form.

The Descriptive Role of Segments: Evidence from Assimilation

ABSTRACT

Francis Nolan
Department of Linguistics
Cambridge University

The phenomenon of assimilation appears, from its traditional segmental description, to involve a discrete change in the phonetic properties of an utterance. Even recent representations within autosegmental formalisms, while capturing the fact that assimilation happens to functional subsets of features, retain the implication that the change is discrete. This paper reports on a line of experimental work which investigates the facts of assimilation of place-of-articulation, using English examples such as red car [red/reg] which contain a potentially assimilable alveolar.

Specifically two main questions were addressed: (i) does articulation mirror the discrete change implied by phonetic and phonological representations of assimilation? And (ii) if assimilation turns out to be a gradual process, how is the articulatory continuum of forms responded to perceptually?

The main tool in the articulatory investigation was electropalatography (EPG). Evidence will be presented which clearly shows that place-assimilation is a gradual process. It is to some extent correlated with increasing rate, but not tied to it. Three analytic categories of 'articulation-type' were imposed on the articulatory continuum: full-alveolar, where alveolar closure was maintained; residual-alveolar, with contact forward along the sides of the palate but no medial closure at the alveolar ridge; and zero-alveolar, indistinguishable from tokens of minimally-distint forms containing no alveolar lexically.

Perceptual identification tests, using stimuli produced using EPG to control for articulation-type, suggested that residual-alveolar forms allowed some identification of alveolars, but zero-alveolar forms were heard the same as non-alveolar minimal pairs. However evidence from an analysis task, in which phoneticians were allowed to listen repeatedly to single stimuli, and from an identification task, where naïve subjects heard minimally contrastive pairs of utterances, suggest that nevertheless cues to the alveolar persist even when there is no evidence in the EPG contact pattern.

The problem of modelling assimilation is then discussed. Phonological representations, even allowing for a notational modification in the autosegmental treatment, are argued to be too discrete. On the other hand to treat place-assimilation as merely an automatic consequence of the physiology and motor control of the vocal apparatus is not satisfactory because there is evidence that it is under speakers' control - not least because its application varies between languages and even varieties of the same language.

It seems therefore that assimilatory behaviour must join the growing list of sub-contrastive phonetic phenomena, such as the timing of voicing and the production of epenthetic stops, which appear to be specified in detail in a speaker's phonetic plan.

Nasal Consonants and The Development of Vowel Nasalization

J. M. Hombert and G. Puech
LAPHOLIA - CRLS
Université Lumière - Lyon 2
69500 Bron FRANCE

It is generally assumed that the development of nasalized vowels is conditioned by certain phonetic factors:

a. the relative positions of the nasal consonant and the vowel

It has been suggested that regressive assimilation, by which a vowel becomes phonetically nasalized under the influence of a following nasal consonant, and later, when that consonant is lost, becomes phonologically nasal (schema 1 below) is a more common source of nasalization than progressive assimilation (shcema 2)

$$1 \quad VN > \tilde{V}N > \tilde{V}$$
$$2 \quad NV > N\tilde{V} > \tilde{V}$$

b. place of articulation of the nasal consonant

According to Chen (1975), nasalization occurs first before labial nasal consonants, second before dental nasal consonants, and finally before velar nasal consonants.

c. vowel quality

Low vowels seems to be affected by nasalization before high vowels, and for equivalent vowel height, front vowels nasalize earlier than back vowels.

It should be emphasized that these "generalizations" are based on very limited data (almost exclusively Romance and Chinese).

Our purpose is to reevaluate these diachronic "universals" by considering new cases - mostly first-hand- of evolutions of nasal consonants and development of vowel nasalization from Bantu languages of Cameroon, Congo and Gabon.

We will pay particular attention to the evolution of prenasalized consonants (voicing and manner of articulation of postnasal C) and to their effect on preceding vowels. It will be shown that carefully collected data from closely related dialects can suggest very likely historical scenario.

These field data will be supported by instrumental data (nasal flow and acoustic measurements).

A number of counter-examples to the diachronic generalizations proposed above have been found with respect to the role of both the place of articulation of the nasal consonant and the quality of the original oral vowel. In other cases, factors previously unmentioned (to our knowledge) in the literature (e.g. vowel length) seem to play an important role in the development of vowel nasalization.

"Targetless" schwa: an articulatory analysis

Cathe Browman and Louis Goldstein

Investigations of reduced (schwa) vowels in English have shown that these vowels take on the acoustic (e.g., Fowler, 1981) and articulatory (e.g., Alfonso and Baer, 1982) properties of neighboring vowels. Analyses of formant trajectories for medial schwa in trisyllabic sequences (Magen, 1989) have shown that F2 moves from a value dominated by the preceding vowel (at onset) to one dominated by the following vowel (at offset). Such data suggest the hypothesis that schwa may be completely unspecified for tongue position. This would also be consistent with the phonological analyses of schwa in French (Anderson, 1982) as an empty nucleus slot. We tested this "targetless schwa" hypothesis by analyzing articulatory data and simulating the results using a computational model of gestural organization.

Using data from the Tokyo X-ray archive, we analyzed /pV1paxV2pax/ utterances produced by a speaker of American English, where V1 and V2 were all possible combinations of /iy,eh,aa,ah,uw/. For V1, V2 and the medial schwa, the horizontal and vertical positions of two of the pellets on the surface of the tongue were measured--those at the middle (M) and rear (R) of the tongue dorsum. For the schwa, these pellets were measured at two points in time--a "tongue" point, corresponding to an inflection point (or flattening) of a tongue pellet's time function during the schwa, and a "lip" point, corresponding to the point of maximum downward displacement of the lower lip during the schwa (approximately the point of maximum lip opening). The "lip" point occurred in the early to mid portion of the acoustic realization of the schwa, while the "tongue" point occurred very late in the acoustic realization of the schwa. For both the "lip" and "tongue" points, we used stepwise linear regression to determine which combination of three predictors--pellet position during V1, pellet position during V2, and an independent schwa factor--best predicted the position of each pellet. For the earlier "lip" point, all three predictors were necessary to get the best prediction, while for the later "tongue" point, the schwa factor and V2 gave the best prediction; V1 did not contribute substantially. Thus, for this speaker, pellet positions for schwa were determined by a combination of an independent schwa component and the following vowel, with the preceding vowel affecting only the earliest portion of the schwa.

Within the linguistic gestural model of Browman & Goldstein (1987), we would expect such a result to be achieved by overlapping the schwa with V2, with the V1 effects a passive consequence of the differing initial conditions for movements out of different preceding vowels. To test this expectation, we used our computational model to generate gestural scores for the above set of utterances. In these gestural scores, the control regimes for the V1 and schwa gestures were contiguous and non-overlapping (with the control regimes for the consonant gestures overlapping them). In addition, the V2 gesture began at the same time as the schwa gesture. Thus, during the acoustic realization of the schwa (approximately), the schwa and V2 gestural control regimes both affected the tongue movements; the schwa relinquished active control during the following consonant, leaving only the V2 tongue gesture active in the next syllable. Gestural scores displaying this organization were input to the task dynamic model (Saltzman, 1986), and the resulting articulatory

trajectories analyzed in the same manner as the X-ray data. The same overall pattern of results was obtained. That is, V1 contributed significantly to the pellet position at the "lips" point of the simulated data, but not to the pellet position of the "tongue" point. Thus, these aspects of the X-ray data for schwa were adequately modeled by a simulation in which gestural control regimes for V2 and schwa overlapped, but V1 effects on schwa were passively achieved.

Note that an independent schwa component was required in the regression, indicating that the strong form of the "targetless schwa" hypothesis does not adequately characterize these data. In addition, we simulated two approaches to a targetless schwa, one in which only V2 was active during the (acoustic) schwa, and one in which there was no active tongue control of any kind during the initial portion of the schwa. Neither approach adequately modeled the X-ray data. However, in one respect the schwa is targetless, or inherently colorless: its value is equal to the mean of the pellet positions for all the flanking vowels. Thus, the schwa has a "target" value that is completely predicted from the targets for the full vowels.

REFERENCES

Alfonso, P. & T. Baer (1982). Dynamics of vowel articulation. Language and Speech, 25, 151-173.

Anderson, S. (1982). The analysis of French schwa. Language, 58, 535-573.

Browman, C.P., & L. Goldstein (1987). Tiers in Articulatory Phonology, with some implications for casual speech. Haskins Laboratories Status Report on Speech Research, SR-92, 1-30. To appear in J. Kingston & M.E. Beckman (eds.), Papers in Laboratory Phonology I: Between the Grammar and the Physics of Speech. Cambridge: Cambridge University Press.

Fowler, C.A. (1981). Perception and production of coarticulation among stressed and unstressed vowels. Journal of Speech and Hearing Research, 24, 127-139.

Magen, H. (1989). An acoustic study of vowel-to-vowel coarticulation in English. PhD. dissertation, Yale University, Department of Linguistics.

Saltzman, E. (1986). Task dynamic coordination of the speech articulators: A preliminary model. In H. Heuer & C. Fromm (eds.), Generation and modulation of action patterns (Experimental Brain Research Series 15). New York: Springer-Verlag.

Coarticulation Increases With Rate ?

Nigel Hewlett and Linda Shockey

ABSTRACT

In a pilot experiment, we investigated the questions 1) is there a difference
in DEGREE of coarticulation between CV utterances produced in very careful
citation form and CV utterances embedded in connected speech? and 2) is there
a difference in TYPE of coarticulation in these two conditions? We
hypothesised that right-to-left coarticulation might be stronger in citation
form (i.e. the vowels would in some sense be more salient) and that left-
to-right coarticulation might be more in evidence in connected speech, thereby
maximising consonantal features.

The present experiment, based on the speech of one male speaker of Standard
Southern British English, suggests that, contrary to our expectations,
CV sequences in the two speech styles are actually quite different, but that
the major difference lies in an overall lowering of the most prominent burst
frequency in connected speech. Further research will focus on determining
how general this tendency is.

SUNDAY MORNING SESSION

DOWNSTEP IN DUTCH: IMPLICATIONS FOR A MODEL

Rob van den Berg, Carlos Gussenhoven, Toni Rietveld

The research reported attempts to identify the parameters that must be
included in an implementation model for Dutch intonation. The emphasis
so far has been on the issues of phonological representation and phonetic
implementation of descending intonation contours. Two specific questions
about Downstep in Dutch were addressed.

The first concerned Reset, the interruption of Downstep. It is argued
that in Dutch, Reset should be modelled as the composite effect of
Phrasal Downstep, that is a downward shift of the register relative to a
preceding (potentially multi-accented) phrase, and Accentual Downstep,
that is the downward shift of a H* relative to a preceding H* inside a
phrase. The model we propose incorporates this view of Reset and
amounts to a reorganization of the model proposed by Ladd (1987).

The second part studies Accentual Downstep in some more detail by
attempting to model production data from two male and two female
speakers. It particularly addresses the question whether the data can be
fitted with an Accentual Downstep factor that is independent of speaker,
range, and number of downstepped accents. It would appear that this is
indeed the case, although the downstep factor showed a slight dependence
on number of accents. We also present a number of measurements
pertaining to Phrasal Downstep.

Phrasing, prominence and tone in Japanese

Elisabeth Selkirk and Koichi Tateishi
University of Massachusetts/Amherst

Studies of the tonal phonology of Japanese sentences and of the phonetic implementation of tone (McCawley, Haraguchi, Kohno, Poser, Kubozono, Pierrehumbert and Beckman, Selkirk and Tateishi) have assumed, in addition to prosodic structure, a role for the representation of prominence, in that the accents of focussed words are realized higher than those of unfocussed ones in comparable contexts. Phonetic studies by Kubozono and Selkirk-Tateishi provide further evidence for a representation of prominence independent of tone and phrasing.

Instrumental investigations of Major Phrases containing three or more accented words show two things. First, as Pierrehumbert and Beckman report, there appears to be successive downstepping in noun phrases with a right-branching [A [A N]] structure. Yet, as Selkirk and Tateishi (1988a) show, left-branching NP's with the structure [N N] N] show no such succssion of downsteps. That this difference may be ascribable to a difference in Major Phrase-internal prominence relations is made more plausible by a second result on long Major Phrases. Major Phrases consisting of four or five accented words in a left-branching structure show a marked downstep after the first accent, as in the earlier cases, and then, within the downstepped range, an alternating succession of higher and lower peaks on the accents, one which is reminiscent of rhythmic alternating patterns (Kubozono, Selkirk and Tateishi (1988a). Both results suggest that downstepping apples only once in the Major Phrase, after the first accent, and that subsequent pitch values are determined by the relative prominence of the accented words.

The emerging role for a representation of prominence as a factor in the realization of tone in Japanese requires us to consider seriously an alternative to the theory that the phonology and phonetics of Japanese tone are determined by a representation of both phrasing and prominence. The altern-ative is one which does away with phrasing altogether, amd relies solely on the representation of prominence. A number of arguments against the prominence-only theory will be presented, including one based on data from the studies of the distribution of initial lowering in long Major Phrases con-taining unaccented words (Selkirk and Tateishi 1988b).

On the hierarchy of intonational representation:
Evidence from Japanese

Haruo Kubozono
Nanzan University, Japan

Abstract

This paper reports on computer-aided experiments carried out
to explore the linguistic structure of Japanese downstep.
Evidence from these experiments shows that the intonational
process of downstep involves not only a phonological aspect,
which Poser (1984) successfully explored, but also a metrical
aspect which is determined by the surface syntactic
constituency of utterances. The notion of 'metrical boost'
is proposed to account for the systematic effect of syntax on
the configuration of downstep. To be more specific, it is
shown that the metrical aspect of downstep is attributable to
a rather general upstep mechanism triggered by syntax, which
plays the role of modifying the phonologically defined down-
step pattern by raising it in the marked right-branching
structure.

   This upstep principle of metrical boost provides not only
plausible accounts of the intonational differences shown by
syntactically complex utterances but also some significant
insights into the organization of intonational representation
in Japanese. Particularly important is the implication that
intonational representation is hierarchically organized in
much the same way as syntactic representation, an
implication which is inextricably associated with the
question of intonational categories. This implication is
discussed as evidence against the conventional assumption
that intonational representation involves an n-ary branching
structure at the level where downstep is defined. It is
argued instead that it constitutes evidence for the
recursive model proposed by Ladd (1986), in which the
distinction between right-branching and left-branching
structures is made possible by a binary branching recursive
mechanism.

References
Beckman, M. and J.B. Pierrehumbert 1986. Intonational structure
    in English and Japanese. Phonology Yearbook 3. 255-309.
Kubozono, H. 1988. The Organization of Japanese Prosody.
    Ph.D. Thesis, University of Edinburgh.
Kubozono, H. in press. Syntactic and rhythmic effects on
    downstep in Japanese. To appear in Phonology 6-1.
Ladd, D.R. 1986. Intonational phrasing: the case for recursive
    prosodic structure. Phonology Yearbook 3. 311-340.
Poser, W. J. 1984. The Phonetics and Phonology of Tone and
    Intonation in Japanese. Ph.D. Dissertation, MIT.

Janet Pierrehumbert & David Talkin
Linguistics Research Dept.
AT&T Bell Laboratories

SUNDAY AFTERNOON SESSION

In this paper, we examine lenition as the effect of prosodic structure on how segments are pronounced. The segments selected for study are /h/ and glottal stop. These segments permit us to concentrate on the source component of lenition. Although previous studies of lenition have concentrated on vocal tract characteristics, source characteristics are typically also at issue. For example, unstressed voiceless stops are less aspirated than stressed ones, and unstressed vowels are more likely than stressed ones to become devoiced in the vicinity of a spread glottis consonant. Thus, we hope that a study of the laryngeal consonants will result in methods and insights which have broader applicability.

/H/ and GLOTTAL STOP

/h/ and glottal stop are both produced by a laryngeal gesture. They make no demands on the vocal tract configuration, which is therefore determined by the adjacent segments. They are both less sonorous than vowels, because both involve a gesture which reduces the strength of voicing. But the gestures for /h/ and for glottal stop are of opposite polarity; for /h/, the folds are abducted (spread apart), while for glottal stop they are adducted (pressed together). Comparing the lenition of /h/ and glottal stop thus provides an opportunity to abstract the magnitude of the gesture away from its direction. /h/ and glottal stop also differ in their phonological status. /h/ appears as a phoneme in lexical representations and, in the intervocalic context of the experiment, is not thought to exhibit phonological alternation. Thus all observed variation will be attributed to the phonetics. The glottal stop is not distinctive in lexical representations of English, but is rather thought to be introduced by rule to supply an onset for syllables which lack one. (We do not consider glottal stops arising as an allophone of /t/, because we are considering intervocalic consonants only). In analyzing the data, it is thus necessary to consider both whether the rule has applied, and what the phonetic interpretation of the result is.

METHODS

Experimental materials and phonetic measures have been developed together in order to achieve the most interpretable results. The experimental materials consist of meaningful sentences which vary the word and phrase-level prosody. /h/ occurs word-initially and medially, before stressed and unstressed vowels, in nuclear position, post-nuclear position, nuclear position with emphasis, and nuclear position following an intonational phrase boundary. The glottal set lacks word-medial exemplars because of the constraints of American English phonology. Near minimal comparisons among the /h/'s and glottal stops were found with the assistance of computer searches of the lexicon; the segmental frame for the target consonant is always mV_VC. All intonation patterns used had a L tone on the target syllable (and therefore a low f0) because maximal separation between f0 and f1 makes estimation of the source characteristics more reliable.

Acoustic measures are designed to reflect the strength of articulation for the two consonants. For /h/, subglottal zeroes and noise excitation make inverse filtering unreliable, so we use measures that can be computed on the original speech. Analysis is done pitch-synchronously using the output of the Talkin epoch finder. The amplitude ratio of the first two harmonics is used as an index of abduction, and a measure of noise excitation has been developed based on the distribution of energy in the period. We also consider the rms energy during the first portion of the glottal pulse (an indicator of how effectively the formants are excited) and the extent of the segmental effect on f0. The characterization of glottalization has proved to be more difficult. Measures now under study include the spectral balance in the result of inverse-filtering, and the extent of irregularity in the periodicity of the excitation. The time course of these acoustic measures provides a basis for study of the durational structure.

RESULTS

Pilot results indicate that all prosodic variables manipulated affected both the duration and the magnitude of the consonantal gesture. This shows that the pronunciation rules for individual segments have access to all levels prosodic structure; the prosodic structure is as important for the realization of segments as we know it to be for the realization of tones. Boundaries may affect the duration disproportionately, as compared to stress. The /h/ variation is readily viewed as gradient. On the other hand, there is no trace of a glottal stop in many utterances in the glottal stop set, particularly when the target syllable is both weak and in a weak prosodic position. Study of the distribution suggests that the glottal stop insertion rule is sensitive to phrase-level prosody, not just word and syllable structure. Data from the experiment is still being analyzed, and we hope to present more definitive results at the meeting.

Prosodic structure and tempo
in a sonority model of articulatory dynamics

Mary E. Beckman   Jan Edwards   Janet Fletcher
Ohio State University   Hunter College, CUNY   Ohio State University

The measured acoustic durations of segments are considerably longer in some prosodic contexts than in others. The characterization of these lengthenings and of their relationship to tempo variation and coarticulation has posed many problems for models of speech production. A description in terms of dynamic parameters such as gestural stiffness promises to resolve some of these problems by providing a more explanatory phonetic representation of articulations (see, e.g., Kelso et al. 1985, Saltzman 1986). Specifying sequences of segments as an orchestration of such dynamic specifications has provided an illuminating interpretation of autosegmental representation within the syllable (Browman & Goldstein 1987, 1988, this volume). In this paper, we consider how such a task-dynamic model can account for the durational correlates of intonation-phrase boundaries and accentual prominence. We have analyzed the peak velocities, displacements, and durations of opening and closing mandibular gestures for [pap] sequences in final versus non-final position and in nuclear-accented versus unaccented syllables spoken at three different tempi by four subjects. In general, the kinematics of the slow/normal tempo contrast are also typical of the final/non-final contrast. Both final syllables and syllables spoken at a slow tempo are associated with slower gestures without any marked increase in gestural displacement, although, unlike the overall tempo change, final lengthening slows the closing gesture much more than it does the opening gesture. In the task-dynamic model, these results can be described by specifying a lesser gestural stiffness, an increase of the gesture's "settling time" (Saltzman & Munhall, in press). The durational increase of nuclear accent, on the other hand, is quite different. Nuclear accent is associated with no consistent velocity decrease. Indeed, nuclear-accented opening gestures are faster than unaccented opening gestures. Accented gestures are also consistently more displaced. In a task-dynamic model, the greater velocity would result from specifying such a greater displacement with no change in gestural stiffness. These results suggest that the durational effect of accent must come from specifying a longer "activation time" (Saltzman & Munhall, in press) for the opening gesture; the syllable is longer because the consonant gesture begins later relative to the vowel. The acoustic effect of this phase specification would be to increase the period of maximum jaw opening, increasing the total energy in the accented vowel and making it more sonorous. Final lengthening, by contrast, would be functionally equivalent to a local tempo change, a slowing down that reaches in from the phrase edge. There is only one difficulty with this interpretation. For at least one of the subjects, the very longest syllable type (final accented pop at slow tempo) showed a similar peak velocity to the contrasting non-final syllable type, even though the syllable's overall duration was just as much longer than that of its non-final counterpart as at the normal tempo. Thus, when a limit on velocity decrease is reached, intrasyllabic activation times seem to be reorganized so that the targeted lengthening can be maintained by a relatively later closing gesture. This suggests that the prosodic effects are specified in terms of a complex phonetic function that relates dynamic specifications to a more abstract representation of sonority over time.

Modelling assimilation in non-segmental rule-free synthesis [1]

John Local

Experimental Phonetics Laboratory
Department of Language and Linguistic Science
University of York
Heslington
York YO1 5DD

Abstract

Synthesis-by-rule implementations have shown remarkably little interest in
employing coherent phonological models. Where such systems have used
'phonological' rules this has usually been in a piecemeal and ad hoc fashion.
Moreover the kinds of phonological 'rules' which have been adopted in
speech synthesis have typically been of a segmental Generative Phonology
rewrite-kind. In this paper I will demonstrate that significant benefits can
be gained from employing a coherent phonological approach in synthesis.
The phonological approach I will describe is one which (a) does not admit
of segmental entities at any level of representation, (b) distinguishes strictly
between phonology and phonetics, and (c) does not admit of rewrite rules.
The modelling of 'process' phenomena would seem to raise serious
problems for this approach. It does not. Taking 'assimilation' effects in
British English as an example I will show that the phonological approach
proposed makes it possible to model them in a non-arbitrary and
naturalistic fashion.

Abstracts of the other two papers in the Monday morning session, by
Loe Boves and by Brian Pickering and John Kelly, will be available
at registration on 29 June.