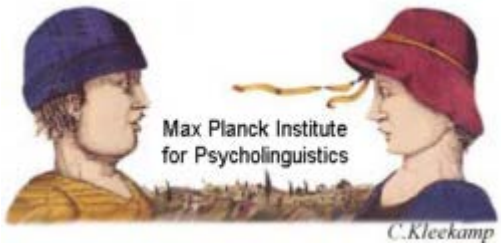


This is an archive site

LabPhon 7

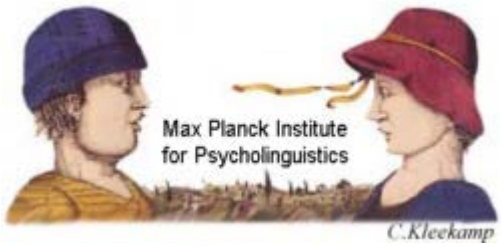
Seventh Conference on Laboratory Phonology
Thursday 29 June - Saturday 1 July 2000

Programme
Call for Papers
Abstracts
Publication
Sponsors
Board

- Hosted by:
  - University of Nijmegen (KUN)
  - Max Planck Institute for Psycholinguistics (MPI)
- Location:
  Collegezalencomplex
  Mercatorpad 1
  University of Nijmegen
  The Netherlands
- Themes and speakers:
  - Phonological encoding
    - Willem Levelt, discussant
      Max Planck Institute for Psycholinguistics
    - Pat Keating, invited speaker
      University of California, Los Angeles
  - Phonological processing
    - Anne Cutler, discussant
      Max Planck Institute for Psycholinguistics
    - Janet Pierrehumbert, invited speaker
      Northwestern University
  - Field work and phonological theory
    - Leo Wetzels, discussant
      Free University of Amsterdam
    - Didier Demolin, invited speaker
      Free University of Brussels
  - Speech technology and phonological theory
    - Louis Boves,discussant
      University of Nijmegen
    - Aditi Lahiri, invited speaker
      University of Konstanz
  - Phonology-phonetics interface
    - Bruce Hayes, discussant
      University of California, Los Angeles
    - Nick Clements, invited speaker
      CNRS, Paris
    - John Ohala, invited speaker
      University of California, Berkeley
- Important dates:
  - 14 January 2000
    Deadline for receipt of abstracts
  - 1 March 2000
    Notification of accceptance

- 28 April 2000
  Deadline for receipt of draft papers
- 1 June 2000
  Deadline for excursion sign-up
  Deadline for receipt of advance registration payment
- 29 June 2000
  Conference begins
- 1 December 2000
  Deadline for receipt of final papers

- Organizing committee:
  - Carlos Gussenhoven, KUN
  - Toni Rietveld, KUN
  - Natasha Warner, MPI
- Contact information:
  Please note: If you have been using the email address labphon7@hooft.let.kun.nl, please do not use it anymore. Use the address below instead. This is due to a change in email servers.
  - Email
    labphon7@let.kun.nl
  - Fax
    +31 (0)24 352 1213
  - Post
    LabPhon 7
    Max Planck Institute for Psycholinguistics
    Postbus 310
    6500 AH Nijmegen
    The Netherlands



LabPhon 7

Themes

Phonological encoding

Once speakers have decided what words to use, they will have to retrieve the corresponding phonological forms from their lexicons, build phonological representations of utterances, and pass this information on to the phonetic implementation module. In some cases, the phonological form of the word will have to be pulled together from different morphemes, as in the case of regular inflections in English, while for simplex words and irregular forms like "went" (GO-Past) no such assembly is needed. Where does the cut-off point lie here? For instance, are null-inflections assembled? Can speakers retrieve phonological forms "surgically", without interference from phonological forms of semantically related words? And what is the form of this phonological representation: are segments syllabified and syllables metrified, or is the full representation still to be assembled? Are the segments fully specified or will redundant features have to be supplied, and if so at what stage? And once a full surface representation has been constructed, complete with higher prosodic structure, what sort of chunks are passed on to the phonetic implementation? Are articulatory programmes built from scratch, starting from the feature, or can ready-made higher-level programmes be called on? And what happens when such ready-made programmes cross word-boundaries and disturb the word-based prosodic structures computed or retrieved earlier?

Phonological processing

The information available to hearers in the acoustic speech signal is very large indeed. When and how do listeners use what in order to decide what they hear? A considerable research effort has been devoted to the question of how listeners decide where words begin, that is, when it is useful to entertain the hypothesis that a certain part of the speech stream constitutes the beginning of a word. Listener strategies appear to be language-specific: only if one's language has word-based vowel harmony does it make sense to take sequential vowel qualities into consideration when deciding where to begin a lexical search. Likewise, the extent to which speakers may provide strings of perceived segments with syllable structure may depend on the role which syllable structure plays in the phonology of the language. How do speakers cope with context-dependent segmental insertions, deletions and assimilations? What role is played by the prosodic and tonal structure? What kind of representation is available in the lexicon with which perceived phonetic or phonological features can be matched during a search? And when and how are various types of non-phonological information brought to bear on the recognition process?

Field work and phonological theory

Phonological theory aims to account for the shapes of the sound systems of the world's languages. What segments, what metrical, tonal and prosodic structures do languages have, how do they combine linearly and hierarchically, and why are these segments and structures statistically distributed the way they are, within and across languages? A prerequisite to serving this aim is the availability of reliable data. Although it may go too far to say that every new language provides at least one aspect which overturns our unspoken conceptions of what can exist, it is fair to say that new data continue to throw unexpected light on current conceptions of phonological structure, and we are far from feeling confident that we know what we need to know. The current threat of extinction that looms over the stock of spoken languages makes the crucial role of field work all the more conspicuous. Greater mobility, together with the availability of high-quality recording and analysis techniques, have widened our notion of "field" in the sense that the field may be the laboratory at the other end of the corridor from our offices, and that the crucial

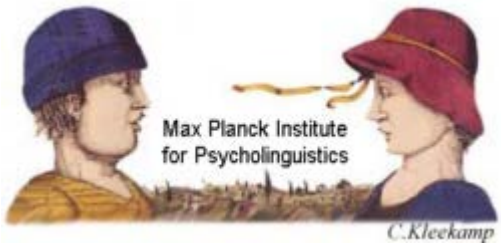
element in this theme is theoretical advances built on "primary data".

Speech technology and phonological theory

More so than in speech synthesis, advances in speech recognition have been possible without any appreciable contribution of phonological theory. This is because researchers have worked with pattern recognition techniques which are independent of the medium within which those patterns exist. Indeed, success in optical recognition has likewise been possible in the absence of a theory of visual perception. In many ways, however, the success of current speech recognition systems is limited. Personalised dictation systems as well as systems with unlimited numbers of speakers and limited sets of expressions to be recognised fall far short of the performance achieved by humans, who recognise unlimited sets of expressions spoken by unlimited sets of speakers. A breakthrough can perhaps be forced by a consideration of the way humans identify linguistic units in a situation where their acoustic properties are highly variable, due to interactions with speaking style, speaker, and nearby other units. This would allow the current knowledge-shy recognition strategy to be replaced with one that makes non-trivial use of phonological representations. Possibly, too, progress in speech synthesis can be based on a careful implementation of phonological accounts, particularly in the area of prosody.

Phonology-phonetics interface

Phonological representations are conglomerates of discrete features and structures selected from a finite set. This is how we assume humans solve the onerous task of remembering the sound forms of the words they know. What speakers produce and hearers receive, however, are continuously varying acoustic patterns, whose shapes are determined by the ergonomics of vocal sound production and perception. It is evident that the nature of the set of features and structures is historically indebted to the ergonomics of speaking and perceiving, and that many changes that occur over time are at least in part determined by these ergonomics. But how direct are these relations? Does a speaker's phonology change with every change in the phonetics, and if not, how much leverage are speakers allowed in the phonetic implementation? Or is the notion that phonological representations and phonetic implementation are separate modules perhaps misguided? How phonetic are phonological representations, and to what extent is phonetic implementation aiding and abetting in the signalling of phonological contrasts? Do features refer to articulatory states, to articulatory gestures, or to auditory effects, or perhaps to all of these at the same time?



LabPhon 7

Seventh Conference on Laboratory Phonology
Thursday 29 June - Saturday 1 July 2000

Thursday, 29 June 2000

8:00 Registration and coffee
Collegezalencomplex (CC)

8:50 Welcome!
Lecture Room Building, Room 1 (CC1)

Session 1: Phonological Encoding

9:05 Introduction
Pim Levelt (Max Planck Institute for Psycholinguistics)

9:15 Invited talk
A phonetician's view of phonological encoding
Pat Keating (University of California, Los Angeles)

10:00 Phonological variation as evidence for lexical representation of homonyms
Daniel Jurafsky, Alan Bell, and Cynthia Girand (University of Colorado, Boulder)

10:30 Coffee

11:00 Phonological encoding: In search of the lost syllable
Niels Schiller (Max Planck Institute for Psycholinguistics), Albert Costa (Harvard University), and Angels Colomé (University of Barcelona)

11:30 Temporal distribution of interrogativity markers in Dutch: A perceptual study
Vincent van Heuven (Leiden University) and Judith Haan (University of Nijmegen)

12:00 Comments
Pim Levelt (Max Planck Institute for Psycholinguistics)

12:20 Lunch
De Refter

Session 2: Phonological Processing

13:50 Introduction
Anne Cutler (Max Planck Institute for Psycholinguistics)

14:00 Invited talk
Morpheme boundaries, word boundaries, and glottal stops
Janet Pierrehumbert (Northwestern University)

14:45 Phoneme frequency in spoken word recognition
Danny Moates, Z.S. Bond, and Verna Stockmal (Ohio University, Athens)

15:15 Coffee

15:45 Visual cues in the perception of speech timing and their implications for phonological structure
Haruo Kubozono (Kobe University)

16:15 A typological study of stress 'deafness'
Sharon Peperkamp and Emmanuel Dupoux (Laboratoire de Sciences Cognitives et Psycholinguistique, Paris)

16:45 Contextual variability, speaking style, and language background
Ann Bradlow (Northwestern University)

17:15 Comments
Anne Cutler (Max Planck Institute for Psycholinguistics)

17:35 Close

Friday, 30 June 2000

Session 3: Field Work and Phonological Theory

9:05 Introduction
Leo Wetzels (Free University Amsterdam)

9:15 Invited talk
The search for primitives in phonology and the explanation of sound patterns: The contribution of fieldwork studies
Didier Demolin (Free University Brussels)

10:00 Acoustic correlates of rhythm classes
Esther Grabe (University of Cambridge) and Ee Ling Low (Nanyang Technological University, Singapore)

10:30 From pitch accent to stress accent in Basque and the typology of accentual systems
José I. Hualde (University of Illinois, Urbana) Gorka Elordieta, Gaminde Inaki (University of the Basque Country), and Raika Smiljanic' (University of Illinois, Urbana)

11:00 Coffee

Poster Session

11:30 Poster session (even)
Main Library Art Gallery

12:30 Lunch
De Refter

13:30 Poster session (odd)
Main Library Art Gallery

Session 3 continued

14:30 Predicting the relative importance of prosodic cues from linguistic structure: The case of stress and tone in the word prosody of Samate Ma`ya
Bert Remijsen (Leiden University)

15:00 Comments
Leo Wetzels (Free University Amsterdam)

15:20 Tea

Session 4: Speech Technology and Phonological Theory

15:50 Introduction
Louis Boves (University of Nijmegen)

16:00 Invited talk
Underspecified recognition
Aditi Lahiri (University of Konstanz)

16:45 Phonetic features in ASR: A linguistic solution to acoustic variation?
Jacques Koreman, Bistra Andreeva and William Barry (University of the Saarland)

17:15 Comments
Louis Boves (University of Nijmegen)

17:35 Close

Dinner and excursion

Saturday, 1 July 2000

Session 5: Phonology-phonetics Interface

9:05 Introduction

Bruce Hayes (University of California, Los Angeles)

9:15 Invited talk
Explosives, implosives, and nonplosives: Some linguistic effects of air pressure differences in stops
Nick Clements (CNRS, Paris) and Sylvester Osu (LLACAN-CNRS, Villejuif)

10:00 Assimilatory processes and aerodynamic factors
Maria Josep Solé (Autonomous University of Barcelona)

10:30 Coffee

11:00 Gestural overlap and recoverability: Articulatory evidence from Georgian
Ioana Chitoran (Dartmouth College), Louis Goldstein (Yale University and Haskins Laboratories), and Dani Byrd (University of Southern California)

11:30 Tonal association and target alignment: Implications for intonation theory
Sónia Frota (University of Lisbon)

12:00 Lunch
De Refter

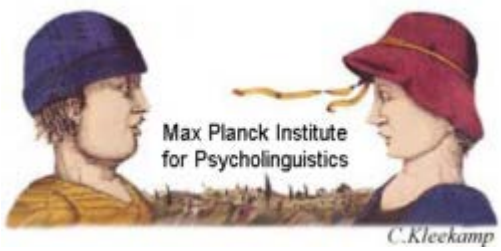
Session 5 continued

13:30 Invited talk
The phonetics of phonologization
John Ohala (University of California, Berkeley)

14:15 Comments
Bruce Hayes (University of California, Los Angeles)

14:35 Close

Katholieke Universiteit Nijmegen



Call for Papers

LabPhon 7

Seventh Conference on Laboratory Phonology

Nijmegen, The Netherlands, 29 June - 1 July 2000

LabPhon 7 will center around five themes, each with at least one invited speaker and discussant:

Phonological processing
Janet Pierrehumbert, invited speaker
Northwestern University
Anne Cutler, discussant
Max Planck Institute for Psycholinguistics

Speech technology and phonological theory
Aditi Lahiri, invited speaker
University of Konstanz
Louis Boves, discussant
University of Nijmegen

Phonological encoding
Pat Keating, invited speaker
University of California, Los Angeles
Willem Levelt, discussant
Max Planck Institute for Psycholinguistics

Phonology-phonetics interface
Nick Clements, invited speaker
CNRS, Paris
John Ohala, invited speaker
University of California, Berkeley
Bruce Hayes, discussant
University of California, Los Angeles

Field work and phonological theory
Didier Demolin, invited speaker
Free University of Brussels
Leo Wetzels, discussant
Free University of Amsterdam

We invite submission of abstracts relating to any of these five themes. Please send a single copy of your abstract either by postal mail to

LabPhon 7
MPI
Postbus 310
NL 6500 AH Nijmegen
The Netherlands

or by email to

labphon7@let.kun.nl

Abstracts must arrive at MPI by 14 January 2000. Please leave sufficient time for delivery.

Abstracts sent by postal mail may contain up to one page of text and up to one additional page for references, figures, examples, etc.

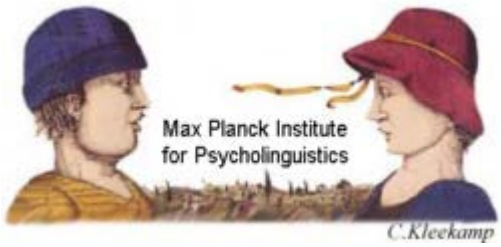
Abstracts submitted by email may contain up to 800 words (including references, examples, etc.) and should be

submitted as the main body of the message (not as any type of attachment). They should contain ASCII text only; for phonetic symbols in emailed abstracts, please use the SAMPA symbols described at http://www.phon.ucl.ac.uk/home/sampa/home.htm.

Faxed abstracts will not be accepted.

Abstracts should be anonymous. If sending an abstract through the post, please attach a card to the abstract stating your name, mailing address, email address, affiliation, title of your abstract, and preference for poster or oral presentation, as well as the theme which you feel your abstract most closely addresses. For email submissions, list this information at the bottom of your message. The organizers reserve the right to assign accepted abstracts to an alternate theme and to assign abstracts to poster presentation.

Authors should note that the schedule for publication for the LabPhon proceedings is being changed as of LabPhon 7. Authors will be required to submit draft papers by 28 April 2000, and to bring nearly final versions of their papers (to be sent out for review) to the conference. The deadline for absolutely final versions (after the review process) will be 1 December 2000. Only papers accepted for oral presentation will be considered for publication.



LabPhon 7

Abstracts

The list of abstracts below is arranged alphabetically by last name of the author or authors. The letter P in the session column indicates a poster, while a digit indicates the session number for a talk. The names in the second column serve as links to the abstracts.

A B C D E F G H I J K L M
N O P Q R S T U V W X Y Z

Session	Name
P	Albano, Eleonora Calvalcante
P	Alter, Kai
4	Andreeva, Bistra
P	Andreeva, Bistra
P	Aoyama, Katsura
P	Arvaniti, Amalia
4	Barry, William
P	Barry, William
1	Bell, Alan
P	Bell, Alan
P	Bell, Alan
2	Bond, Z.S.
2	Bradlow, Ann
5	Byrd, Dani
P	Byrd, Dani
P	Chen, Larissa
5	Chitoran, Ioana
P	Choi, Hansook
5	Clements, Nick
P	Cohn, Abigail
P	Cole, Jennifer
1	Colomé, Angels
P	Corneau, Caroline
1	Costa, Albert
P	Crasborn, Onno

P	Cutler, Anne
P	Delvaux, Véronique
3	Demolin, Didier
P	Demolin, Didier
P	Demolin, Didier
2	Dupoux, Emmanuel
3	Elordieta, Gorka
P	Fitzpatrick, Jennifer
P	Fleischhacker, Heidi
5	Frota, Sónia
P	Frota, Sónia
P	Gilbers, Dicky
1	Girand, Cynthia
P	Girand, Cynthia
5	Goldstein, Louis
P	Goldstein, Louis
3	Grabe, Esther
P	Gregory, Michelle
P	Gussenhoven, Carlos
5	Haan, Judith
P	Haan, Judith
P	Hajek, John
P	Hamann, Silke
P	Hansson, Petra
P	Heijmans, Linda
P	Helsloot, Karijn
5	Heuven, Vincent van
P	Hewlett, Nigel
P	Hruska, Claudia
3	Hualde, José I.
3	Inaki, Gaminde
P	Iskarous, Khalil
P	Janse, Esther
P	Jessen, Michael
P	Johnson, Keith
P	Jongman, Allard
P	Jun, Sun-Ah
1	Jurafsky, Daniel
P	Jurafsky, Daniel

1	Keating, Pat
P	Kim, Hyunsoon
P	Kingston, John
P	Ko, Eon-Suk
P	Kochetov, Alexei
4	Koreman, Jacques
P	Kraehenmann, Astrid
2	Kubozono, Haruo
4	Lahiri, Aditi
P	Lahiri, Aditi
P	Lattner, Sonja
P	Lavoie, Lisa
P	Lavoie, Lisa
3	Low, Ee Ling
P	McDonough, Joyce
P	Miller-Ockhuizen, Amanda
2	Moates, Danny
P	Mohanan, Tara
P	Mücke, Doris
P	Nagao, Kyoko
P	Odé, Cecilia
5	Ohala, John J.
P	Ohala, John J.
P	Ohala, Manjari
P	Olson, Ken
5	Osu, Sylvester
P	Otake, Takashi
2	Peperkamp, Sharon
P	Peters, Joerg
2	Pierrehumbert, Janet
P	Post, Brechtje
P	Pouplier, Marianne
P	Raymond, William
3	Remijsen, Bert
P	Rialland, Annie
P	Rietveld, Toni
P	Schafer, Amy
1	Schiller, Niels
P	Scobbie, James
P	Shattuck-Hufnagel, Stefanie

3	Smiljanic', Raika
P	Smith, Caroline
5	Solé, Maria Josep
P	Soquet, Alain
P	Steinhauer, Karsten
P	Steube, Anita
2	Stockmal, Verna
P	Streefkerk, Barbertje
P	Stuart-Smith, Jane
P	Timmins, Claire
P	Turk, Alice
P	Tweedie, Fiona
P	Vigario, Marina
P	Warner, Natasha
P	Yigezu, Moges
P	Yoneyama, Kiyoko
P	Ziegler, Wolfram

Not Found

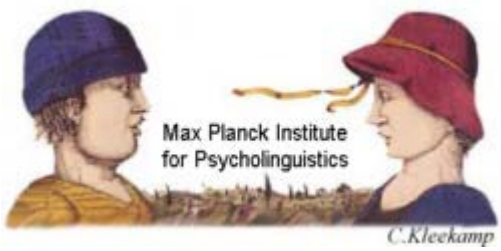
The requested document was not found on this server.

Web Server at labphon7.ruhosting.nl

Katholieke Universiteit Nijmegen

LabPhon 7

Publication



The LabPhon 7 Proceedings will be published in August 2002 as

Gussenhoven, Carlos & Warner, Natasha (eds.)
Laboratory Phonology 7. Berlin/New York: Mouton de Gruyter.

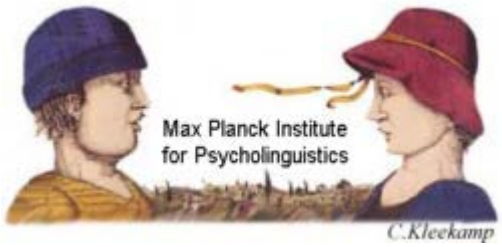
Table of Contents

Acknowledgements		xi
Carlos Gussenhoven & Natasha Warner	Introduction	xiii-xvii
Dan Jurafsky, Alan Bell & Cynthia Girand	The role of the lemma in form variation	1-34
Niels O. Schiller, Albert Costa & Angels Colomé	Phonological encoding in single words	35-59
Vincent J. van Heuven & Judith Haan	Temporal distribution of interrogativity markers in Dutch	61-86
Willem J.M. Levelt	Phonological encoding in speech production: Comments on Jurafsky et al., Schiller et al., and van Heuven et al.	87-99
Janet B. Pierrehumbert	Word-specific phonetics	101-139
Danny R. Moates, Z.S. Bond & Vera Stockmal	Phoneme frequency in spoken word reconstruction	141-169
Haruo Kubozono	Temporal neutralization in Japanese	171-201
Sharon Peperkamp & Emmanuel Dupoux	A typological study of stress 'deafness'	203-240
Ann R. Bradlow	Confluent talker- and listener-oriented forces in clear speech production	241-273
Anne Cutler	Phonological processing: Comments on Pierehumbert, Moates et al., Kubozono, Peperkamp & Dupoux, and Bradlow	275-297
George N. Clements & Sylvester Osu	Explosives, implosives and nonexplosives: The linguistic funcrtion of air pressure differences in stops	299-350
Maria-Josep Solé	Assimilatory processes and aerodynamic factors	351-386
Sónia Frota	Tonal association and target alignment in European Portuguese	387-418
Ioana Chitoran, Louis Goldstein & Dani Byrd	Gestural overlap and recoverability: Articulatory evidence form Georgian	419-447
Bruce Hayes	The Phonetic-Phonology Interface: Comments on Clements & Osu, Solé, Frota, and Chitoran et al.	449-454
Didier Demolin	The search for primitives in phonology and the explanation of sound patterns: The contribution of field work studies	455-513
Esther Grabe & Ee Ling Low	Derivational variability in speech and the Rhythm Class Hypothesis	515-546

José Ignacio Hualde, Gorka Elordieta, Iñaki Gaminde & Rajka Smiljanic'	From pitch accent to stress accent in Basque	547-584
Bert Remijsen	Lexically contratsive stress and lexical tone in Ma'ya	585-614
W. Leo Wetzels	Fieldwork and phonological theory: Comments on Demolin, Grabe & Low, Hualde et al., and Remijsen	615-635
Aditi Lahiri & Henning Reetz	Underspecified recognition	637-675
Dafydd Gibbon	Comments on Lahiri & Reetz	677-685
Subject Index		687-691
Author Index		693-715
Language Index		717-719

---

2 July 2000



LabPhon 7

Sponsors

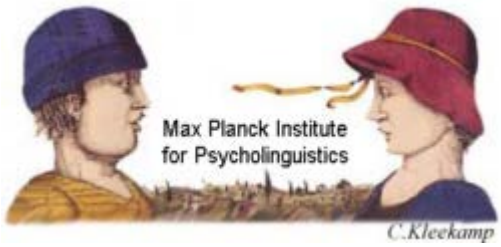
Netherlands Organization for Scientific Research (NWO)

Royal Dutch Academy of Sciences (KNAW)

University of Nijmegen (KUN)

Faculteit der Letteren, University of Nijmegen

Max Planck Institute for Psycholinguistics (MPI)



LabPhon 7

Standing Committee

The ongoing series of LabPhon conferences is guided by a standing committee comprised of past LabPhon organizers. As of the year 2000 the committee consists of the following individuals:

Chair

Amalia Arvaniti (University of Cyprus)

Members

Michael Broe (Ohio State University)
Bruce Connell (Oxford University)
Carlos Gussenhoven (University of Nijmegen)
John Local (University of York)
Richard Ogden (University of York)
Janet Pierrehumbert (Northwestern University)
Toni Rietveld (University of Nijmegen)
Ros Temple (University of York)
Natasha Warner (Max Planck Institute for Psycholinguistics)

Outgoing chair

Pat Keating (University of California, Los Angeles)

Defining the tone system in Mpur: From perception to phonological contrast

Cecilia Odé (Leiden University)

Language

Mpur (West Papuan Phylum) is a Non-Austronesian phylum-level isolate with approximately 5000 speakers in the Northeast Bird's Head Area, Irian Jaya, Indonesia. Three lexical tones have been attested: high, mid and low (Odé 1996, 1998). The analysis of experiments carried out during fieldwork (1998) is still in progress: the issue is whether two more tones are phonologically contrastive: midrising and fall&rise. In this paper type fall&rise (henceforth: type fr) will be discussed.

Introduction

In Odé (1996, 1998) type fr is described as existing on the perceptual level in final syllables of mainly disyllabic words frequently ending in a nasal. Type fr is suspect in the sense that, in contrast to high, mid and low tones, it exclusively occurs word final, that speakers are not consistent in their realization of type fr, and that syllables with type fr ending in a nasal are longer than such syllables not ending in a nasal, or than final syllables with other types of tone. Also, type fr may show two realizations: fall&rise and low/low-falling. With all limitations just described, I did not expect type fr to be phonologically contrastive. Then what is the status of this perceptually so salient type?

Experiment

In order to find out whether type fr is phonologically contrastive and whether its form is a complex contour, I conducted a perception experiment. From a wordlist I selected 33 disyllabic words ending in a nasal (14), a non-nasal consonant (10) and a vowel (9), 19 stimuli with type fr and 14 distractors with other types of tone in the final syllable. Using an analysis-by-resynthesis method, the 19 type fr stimuli were changed into

1. a low/low falling movement, removing the final rise, and
2. a rising movement entirely situated in the final syllable, shifting the timing of the rise.

For the symmetry of the experiment, tones in final syllables of the 14 distractors were also changed into two other types of tone. All 33 originals with two manipulations were included in the experiment.

Reasons for the specific manipulations 1) and 2) of type fr were the following hypotheses:

a. if original and manipulation 1) match, type fr is an allotone of low tone;
b. if original and manipulation 2) match, type fr is probably an allotone of mid or high tone, depending on the level reached.

On the basis of previous observations and the limitations described above, I did not expect that all original type fr stimuli would be the only acceptable realizations and that type fr is a phonologically contrastive complex contour tone. I expected hypothesis a) to be most probable, but I also expected that type fr stimuli with (long) final nasal may be crucial to finding the contrast. In the given context, other hypotheses do not make sense.

The experiment was carried out in the field. Two native speakers were asked to listen to the stimuli, to translate only correct realizations into Indonesian and to reject stimuli that sounded odd to them; a feasible task in my fieldwork experience. Stimuli were twice presented in random order.

Results and conclusion

For type fr stimuli, translations of original and manipulation 2) were not the same or no translation was given for manipulation 2); hypothesis b) is rejected. This was very likely to happen, given the large perceptual difference between type fr and manipulation 2). For type fr stimuli not ending in a nasal (e.g. diti 'to cut', amop 'turtle'), the same translation was given for original and manipulation 1); for these stimuli hypothesis a) is verified. For type fr stimuli ending in a nasal, only original realizations were translated; for these stimuli hypothesis a) is rejected.

I measured final nasals of type fr stimuli and found duration values similar to vowel durations (e.g. firim 'running nose' (-m 220ms)); no reduction of the duration of vowels preceding such nasals was observed. Defining these final nasals as syllabic (initial syllabic nasals exist: n_jep 'firewood'), e.g. fi-ri-m_, type fr realizations in stimuli with final nasal are then a sequence of two tones (low-mid or low-high, depending on the final level reached) in two adjacent syllables V_N.

In my presentation I will discuss more details and considerations that led to my conclusion based on experimentally verified results that could not have been obtained without using the analysis-by-resynthesis method.

References

- Odé, C. 1996. "Mpur tones and intonation in an Amberbaken myth". In Studies in Irian Languages, Ger P. Reesink (ed.). NUSA 40, 61-96.
- Odé, Cecilia.1998. "The bird said 'I am here': a prosodic study of the waimon story in Mpur". In Perspectives on the Bird's Head of Irian Jaya, Indonesia, Jelle Miedema, Cecilia Odé, Rien C. Dam (eds). Amsterdam: Rodopi, 575-602.

Syllable structure and phonological encoding: Evidence from neurolinguistics

Sonja Lattner (Max Planck Institute of Cognitive Neuroscience) and Wolfram Ziegler (City Hospital München-Bogenhausen)

Introduction:

With the development of nonlinear phonology, the syllable has (re-)emerged as an important concept of contemporary phonological and psycholinguistic research. Yet there are still a number of phenomena that cause considerable debate, and even general aspects of the syllable's role in phonological encoding are far from being undisputed [1,2]. One of these controversial issues, the processing of syllable boundaries, will be discussed in the present investigation, where an empirical neurolinguistic approach is taken.

Production errors from speech-impaired adults in a repetition test were analyzed to reveal further evidence about the encoding of syllables in general, with a particular focus on the types of syllable boundaries in German.

Method:

Ten patients (German native speakers) suffering from apraxia of speech (9 of whom were also aphasic) were given a word repetition task.

The material consisted of six German word lists and six lists of phonologically possible nonwords taken from Liepold (1998), which had to be repeated by the subjects word by word. The nonwords were phonologically similar to the lexical items, e.g. "Decke" vs *"Dücke", and both ranged from 2 to 4 syllables.

Altogether, the items contained 1440 word-medial syllable boundaries which were classified by their consonantal structure:

I)	Type I : Cluster (N=230)	
	Heterosyllabic consonants	: C1.C2
II)	Type II : Ambisyllabic structure (N= 380)	
	Ambisyllabic consonants (per definition following the short vowel of an open, stressed syllable)	: X
III)	Type III : Single consonant (N= 830)	
	Single Consonant (following a long vowel)	: (V:).C

In order to investigate whether these target structures vary in their encodability, misproductions ("substitutions") of the boundary consonants were analyzed.

Results:

Tab.1: Percentage of misproduced syllable types

C1C2	X	V:C
27%	9%	17%

Tab.2: Relation of the target structures and the boundary types produced by the subjects. (Percent errors per target structure).

		TARGET STRUCTURE		
		C1C2	X	V:C
REALIZED STRUCTURE	C1C2	93%	25%	12%
	X	2%	67%	1%
	V:C	5%	8%	87%

1. Consonant clusters (C1C2) turned out to be relatively difficult to produce (27% substitution rate), even when

they were separated by a syllable boundary (Tab.1).

2. Single boundary consonants (17% error rate) and particularly the ambisyllabic segments (9% error rate) were affected far less.

3. During the observed segmental substitution, the syllable-boundary structures tended to be preserved.

4. Although the rate of segment substitution in C1C2-type was high, this boundary structure of C1C2 proved to be particularly stable (93% of the misproductions were of the target structure). Single consonant boundaries, too, were rather stable (87% target match), whereas the ambisyllabic structures were preserved in only 63% of the false productions.

5. Additionally, we did a classification of 250 randomly selected German lemmas of the Celex database and found a distribution of 55% C1C2, 13% X and 31% V:C. Thus, there was a high correspondence between stability of syllable structure and frequency of structure type in German.

Conclusion:

In cases of segmental substitutions by apraxic speakers, syllable-boundary types tended to be maintained. This might be seen as confirming the relative independence of segmental and syllabic encoding in the derivational process. The rate of substitution was rather high for C1C2, which indicates that consonantal clustering, even across syllable boundaries, might lead to (motor programming) problems. In spite of these difficulties, the boundary type of C1C2 seems particularly stable, presumably because the relative dominance of this boundary type in the German language has an effect (in addition, such an effect could account for the facts that a misproduced ambisyllabic structure preferably turns into a C1C2, and V:C almost never turns into type X but rather into C1C2). Ambisyllabic structures are less prone to segment substitution in apraxia of speech, but this structure is less stable than other boundary types. Whether this instability comes from the hypothesized fact that ambisyllabic consonants are "fusions" of two single consonants or from the fact that it is a less frequent structure in German cannot be answered yet.

Altogether, the present investigation should be seen as a first exploratory study. A couple of questions remain to be answered (for instance the role of accent or consonant class). Nevertheless, we feel that neurolonguistic investgations into the problem of syllable boundary might make an important contribution to the understanding of phonological encoding.

References

1. Levelt, W.J.M. (1999): A Theory on Lexical Access in Speech Production. Behavioral and Brain Sciences 22:1(1)
2. Schiller, N.O. (1997): The Role of the Syllable in Speech Production. MPI Series in Psycholinguistics 2, Ponsen and Looijen bv, Wageningen
3. Liepold, M. (1998): Ein Beitrag zur Diagnostik von Nachsprechleistungen bei hirngeschädigten Patienten: Entwicklung der Auswertungsmethode eines Nachsprechtests zur Sprechaparaxie-Diagnostik. EKN, München

---

Phonemic entropy and consonant harmony: The relationship between function load and featural contrast in a coronal heavy system

Joyce McDonough (University of Rochester)

This paper reports on the role of functional load in the relationship between place of articulation (p.o.a.) features and neutralizations in the sound system of Navajo. How do the differing functional loads carried by contrastive features within a system affect the sound patterns of a language? The hypothesis put forward is that functional load is a source of entropy in a system. Systems are prone to enhance any property, phonetic or phonological, that build contrasts, and sacrifice those that don't. This hypothesis require a rich level of phonetic description to be encoded into the grammar.

The study is based on a phonetic analysis of the sound structure of Navajo from fieldwork data gathered from native speakers living on the Navajo Nation. In this paper I will discuss an interesting class of cases: the existence of a regressive consonant harmony in the coronal heavy inventory of Navajo.

Recent work has drawn attention to the role of perceptual salience and articulatory ease in shaping the phonotactic constraints and phonological processes that develop in a grammar (Linblom 1986, Manuel 1990, Whalen 1990, Flemming 1995). The Athabaskan languages in general, and Navajo in particular, have coronal heavy inventories. (Labials are lexically very rare and velars often lenite to co-articulated approximants.) Perceptual salience arguably gives advantage to the maintenance of distinctions in a system, while articulatory ease gives advantage to the reduction of articulatory gestures. So why does a coronal heavy grammar like Navajo develop consonant harmony, which results in a considerable loss of p.o.a. distinctions in the lexicon?

One reasonable explanation is found in the functional load carried by p.o.a contrasts in the Navajo sound system. It is arguably light (table 1). Note that the three stop series in the inventory contrast in their aperture features: plain, lateral and fricated releases. Furthermore, each series exhibits a three way laryngeal contrast. Both the series and its laryngeal contrasts are realized in the aperture or release period of the stops. An examination of the phonetic properties of the stops reveals an interesting and relevant pattern: a consistent stop profile arises which enhances the period after the release of the stop.

For instance, speakers are careful to maintain particular characteristics of the articulation of stops, such as their length (generally 150 - 200ms) and a consistent, near 50-50 ratio of closure to release (McDonough and Ladefoged 1992) (figure 1). While duration and closure-release ratios are often considered to be low level phonetic effects, incidental to phonology, in Navajo they play an important role in supporting the phonemic contrasts.

Perceptual salience will give advantage to structures that support the aperture features in Navajo, namely long release periods. Pressure from articulatory ease will reduce the functionally lighter p.o.a. distinctions through consonant harmony, by maintaining a single feature for coronality across the word. It is an argument of this paper that both the phonological process of consonant harmony and the more phonetic aspects of the stop profile are encoded in the grammar.

Two other classes of cases are discussed which support this hypothesis. The first is a cross-generational sound change called 'de-frication' in effect between older monolingual and younger bilingual speakers. This process further decreases p.o.a. distinctions through lenition and reduces a property of stop contrasts (frication in the release period) that carries a light functional load. The second is a problematic set of consonant mutations on stem initial consonants called the Athabaskan 'd-effects'. A spectrographic analysis shows that a prefix 'd' is incorporated into the onset of a fricative initial stem, resulting in the system's canonical closure-release stop profile. An argument is put forth that the d-effect alternations arose through a constraint ranking of incorporation over epenthesis --which operates elsewhere in the grammar-- because of the high functional value associated to the phonetic profile of the stem initial consonants.

This characterization requires the encoding of a level of phonetic detail not usually associated with the specification of contrasts in a language. However, disregarding this information means a loss of insight into the structure of this sound system and the ways in which it developed and is likely to change.

Table 1

Figure 1

References

- Flemming, E. (1995). Auditory Representations in Phonology, Phd Dissertation, UCLA.
- Lindblom, B. (1986). Phonetic universals in vowel systems. Experimental Phonology. J. Ohala and J. Jaeger. Orlando, Academic Press: 13-44.
- Manuel, S. Y. (1990). "The role of contrast in limiting vowel-to-vowel coarticulation in different languages." JASA 88:1286-1298.
- McDonough, J. and P. Ladefoged (1996). The specification of stop contrasts in Navajo. Dam Phonology: HIL Phonology Papers II. N. a. Vogel: 123-142.
- Whalen, D. H. (1990). "Coarticulation is largely planned." Journal of Phonetics, 18:3-35.

Testing licensing by cue: A case of plain-palatalized coronal contrast in Russian

Alexei Kochetov (University of Toronto)

Recent work in phonology revived the interest in phonetic factors as a source of explanation for various phonological patterns (Flemming 1995, Hamilton 1996, Hume et al. 1999, Steriade 1997, 1999, etc.). One of the directions taken in these works is to account for phonemic neutralization, deriving it from acoustic cues (Licensing by Cue: Steriade 1997). In this view phonological contrasts are neutralized in environments that are poor in terms of phonetic information and are licensed in positions that are high on a scale of perceptibility.

This paper tests the hypothesis Licensing by Cue applying it to the distribution of Russian plain-palatalized contrast in coronal stops in two environments. The hypothesis holds that the maintenance of the contrast /t/ vs. /t/ in a given context should correspond to more acoustic information about the contrast in the signal and its better identification by listeners. At the same time, the neutralization of the distinction should be accompanied by fewer cues and lower recognition of the segments. The results of acoustic (6 talkers) and perceptual (12 listeners) experiments presented in the paper support the hypothesis accounting for the neutralization pattern in the first environment, before consonants (across word boundaries). The plain-palatalized contrast before /k/ is more acoustically and perceptually salient than it is in the environments before /n/ or /s/. Substantial spectral and temporal differences between /t/ and /t/ in the former context contribute to the higher identification rate and lower response time. No audible release before the nasal and the strident fricative /s/ result in a substantially higher confusion between the two segments. This correlates with the maintenance of the /t/ vs. /t/ distinction in the former context and its neutralization in the latter position.

The experiments, however, failed to support the hypothesis in the second environment, after vowels (/a/, /u/, and /i/). A strong effect of vowel both in terms of acoustics and perception has no apparent impact neither on the categorical distribution of the contrast, nor on its frequency.

A revised approach, proposed to account for the results, maintains the key assumption of Licensing by Cue that phonetic, primarily perceptual, factors are the main motivation for neutralization, and phonotactic patterns in general. However, it admits that mapping acoustic input to a phonological contrast is by no means direct. It is affected by a number of grammar-internal and external factors. A listener, presented with an acoustic input, recovers a gesture (or a segment) when there is an adequate acoustic information about it, and misses it or confuses with a similar gesture when some information is missing or inconclusive (cf. Browman & Goldstein 1999, Kawasaki 1982, Ohala 1981, 1983, Silverman 1997). The more informative the context is, the higher is the probability that the segment is recovered. Perceptual hierarchies are a direct consequence of this speaker-listener interaction. However, it is argued that the recovered information about the contrast is only a raw input to the grammar, which is processed in a cognitive mode with a direct reference to the phonological form (Gerfen 1999, Hale & Reiss 2000). The grammar induces certain arbitrariness between the signal and its mental representation, generalizing across some phonological and other linguistic domains. As a result, ranking of phonological constraints on gesture combinations in the grammar will follow the general pattern of the perceptual hierarchies, but will inevitably deviate from them in some arbitrary, language-particular ways.

It is concluded that an explanation for neutralization, and phonotactic patterns in general, should not be sought either only in phonology or only in phonetics, but in the interaction of phonetic factors with the phonological grammar.

References

- Browman, Catherine P. and Louis Goldstein. (1999) (to appear). Competing constraints on intergestural coordination and self-organization of phonological structures. In: Bulletin de la Communication Parle. Grenoble.
- Flemming, Edward Stanton. (1995). Auditory representations in phonology. Doctoral dissertation. University of California, Los Angeles.
- Gerfen, Chip. 1999. Obstruent neutralization in Andalusian Spanish: A critical view of licensing by cue. Talk presented at West Coast Conference on Formal Linguistics, April 1999.
- Hale, Mark & Charles Reiss. (2000). Substance abuse and dysfunctionalism: Current trends in phonology. Linguistic Inquiry 31:1.

- Hamilton, Philip James. (1996). Phonetic constraints and markedness in the phonotactics of Australian aboriginal languages. Doctoral thesis. University of Toronto.
- Hume, Elizabeth, Keith Johnson, Misun Seo, and Georgios Tserdanelis. (1999). A cross-linguistic study of stop place perception. In: Proceedings of the 14th International Congress of Phonetic Sciences (ICPhS99).
- Kawasaki, Haruko. (1982). An acoustic basis for universal constraints on sound sequences. Doctoral dissertation. University of California, Berkeley.
- Ohala, J.J. (1981). The listener as a source of sound change. In: C.S. Masek, R.A. Hendrick, & M.F. Miller (Eds.), Papers from Chicago Linguistic Society parasession on language and behaviour. Chicago: CLS. 178-203.
- Ohala, J.J. (1983). The origins of sound patterns in vocal tract constraints. In: P.F. MacNeilage (ed.), The production of speech. New York: Springer Verlag. 189-216.
- Silverman, Daniel. (1997). Phasing and recoverability. New York: Garland.
- Steriade, Donca. (1997). Phonetics in phonology: The case of laryngeal neutralization. Ms., University of California, Los Angeles.
- Steriade, D. (1999). Directional asymmetries in place assimilation: a perceptual account. Paper presented at ICPhS Satellite meeting on Perception in Phonology. San Francisco.

Korean palatalization is a rule of phonetic implementation: Articulatory and acoustic evidence

Hyunsoon Kim (Sogang University, Seoul)

In the framework of Lexical Phonology, Korean Palatalization has been considered to be divided into two: lexical and postlexical palatalization (e.g., Cho and Sells 1991, Kiparsky 1993, Iverson 1993). In lexical palatalization, stem-final stop consonants /t, t^h/ have been proposed to change into their counterpart "postalveolar" affricates /c,c^h/ when followed by derivational suffixes which begin with /(h)i/, as shown in (1a). According to Kiparsky (1993:290), lexical palatalization in Korean is assumed to "spread the features [-anterior] and [+high]" to the stops /t, t^h/ from a following high front vowel and a separate rule of affrication applies to the intermediate stops, deriving the affricates. On the other hand, postlexical palatalization in Korean has been considered to be the process whereby consonants assimilate in place of articulation to the following high fronted vowel /i/ without morphological restriction, as shown in (1b). Thus it has been regarded as a feature-changing rule (e.g., Kiparsky 1993) or a feature-filling rule (e.g., Iverson 1993), depending on whether a target consonant is specified for [-anterior, +high] or not when the rule applies.

However, the present study provides articulatory (i.e., direct palatograms and linguograms taken from twenty subjects) and acoustic (i.e., wide-band spectrograms and LPC spectra) evidence against the two-rule approach to Korean palatalization. Based on the experimental data, this paper proposes the following: (a) there is no lexical palatalization in Korean, and instead Korean has the phonological process of affrication of inserting the feature [+strident] whereby only the manner of articulation of the original stop consonants /t, t^h/ is changed into their counterpart "alveolar" affricates; (b) the frontness of so-called postlexically-palatalized consonants and lexically-derived affricates is not categorical but gradient because of gradual, interpolatory F_2 transition throughout target consonants, and vowel-to-vowel coarticulation across the consonants; (c) therefore, palatalization -- frontness of a consonant before the high fronted vowel /i/ -- occurs only in phonetics as a coarticulatory effect of the following vowel /i/ in Korean, with no division of lexical and postlexical palatalization: and (d) Korean palatalization is a rule of phonetic implementation which can be well captured in terms of articulatory gestures, not in terms of features, in line with Zsiga (1997) who proposes that categorical phonological rules are represented by features, while gradient phonetic ones are represented by gestures.

Articulatory and Acoustic Experiments:

First, in an attempt to see whether "postalveolar" affricates are derived by virtue of lexical palatalization, as in (1a), this paper examines direct palatograms and linguograms of the underlying affricate /c/ and the stop /t/ in the contexts of *a__a* and *a__i*, and those of the lexically-derived affricate [c] in the context of lexical palatalization. The results show that the highest percentage of Korean affricates are alveolars like the stops, and that the lexically-derived affricate [c] is also much more likely to be alveolar than postalveolar, as shown in Table 1, thus confirming Skalická's (1960) study of Korean affricates. Acoustic (i.e., LPC spectra) data of the lexically-derived affricate [c] also shows the characteristics of alveolar consonants (e.g., Stevens 1993) that the average highest-amplitude spectral peaks of the affricates both at the release and 25-35 ms after the release are above the following vowel's fourth formant for each speaker in the word /mat+i/ 'first child.'

Second, acoustic (i.e., wide-band spectrograms) data shows that the frontness of the lexically-derived affricates and postlexically-palatalized coronal consonants is not categorical, but gradient with inter- and intra-speaker variation. As shown in Figure 1, across the intervening lexically-derived affricate [c] in *mat+i* 'first child' and underlying coronal consonants /t, s, c/ in *a__i* context, the movement of the tongue body, which is acoustically correlated with F_2 transition (e.g., Halle 1959, Fant 1960), is gradual throughout the target consonants. In addition, the preceding vowel /a/ is fronted as the result of the coarticulatory effect of the following high fronted vowel /i/. Another acoustic data of the consonants /t, s, c, n, l/ in the context of *a__i* also shows gradual, interpolatory F_2 transition across the target consonants, and that the consonants are more fronted word-initially than word-medially before /i/.

Data 1

Table 1 Distribution of the five palatographic classifications for 20 Korean speakers for /t, c/ and the lexically-derived affricates /c, c^h/ (percent of all tokens). Modal values are shown in boldface.

	underlying				lexically-derived
Place of articulation	t in *a__a*,	t in *a__i*,	c in *a__a*,	c in *a__i*	c
dental	__	__	__	__	__
dentalveolar	20	11.2	__	__	__
alveolar	**75**	**76.6**	**85**	**94**	**94**
palato-alveolar	5	11.2	15	6	6
alveolo-palatal	__	__	__	__	__

Figure 1

Investigating prosodic structure by means of spoken corpora annotation

Amalia Arvaniti (University of Cyprus)

This paper deals with the role that annotated speech corpora can play in clarifying our understanding of prosody, in particular the relation between the phenomena that prosodic structure purports to represent, phrasing and relative prominence, and those it is expected to "control", sandhi and intonation. It argues that "fieldwork" -- specifically in the form of annotated speech corpora -- can play an important role in clarifying our understanding of the phenomena and therefore in developing the theory. The reason for this is that corpora can provide us with instrumentally (rather than impressionistically) inspected and carefully annotated naturally occurring speech; the study of such data is essential for the understanding of prosodic phenomena, particularly those less amenable to laboratory-type experimentation, such as sandhi. More concretely, the paper discusses the relation between phrasing, stress and sandhi in Greek as this emerges from the inspection of the utterances collected for the development of Greek ToBI (GRToBI).

Although ToBI annotated corpora have been mostly used for research on intonational phonology, their Break Index tier marks prosodic phrasing, and with appropriate additional annotation can provide data on the relation between stress, sandhi and phrasing. Such a procedure has been adopted in GRToBI, which marks stress in its "Prosodic Words" Tier, and also flags with the diacritic "s" all breaks that show evidence of sandhi.

Investigation of the GRToBI database has brought up interesting results about previously reported and newly observed sandhi rules and their domain spans. These tend to be larger than previously reported, thus questioning an assumption underlying much work on prosody (especially within the ToBI framework), namely that sandhi signals cohesion between constituents. More importantly, however, the data provide evidence questioning the rigid stratification of prosodic constituents that is explicitly or implicitly assumed in most models of prosodic structure. Specifically, the data show that (a) the domains of stress and sandhi rules are not always isomorphic (e.g. although enclitics count as part of the prosodic word for stress and intonation purposes, Nasal Deletion and Stop Voicing [which apply within prosodic words, according to Nespor & Vogel, 1986] do not apply between enclitics and their hosts); (b) particular sandhi rules may apply to larger domains yet fail to apply to smaller ones (e.g. Nasal Deletion and Stop Voicing regularly apply across prosodic words in the GRToBI corpus).

Accounting for these sandhi data under current assumptions is problematic. Extending the domain of application of most sandhi rules to the intonational phrase level largely defeats the purpose of using phrasing to account for sandhi. On the other hand, reanalyzing particular strings as smaller domains on the basis of sandhi evidence runs counter to the stress and intonation information; e.g. items that under such an analysis would be considered a single prosodic word retain their stress pattern and accent (but Greek words have only one stress). Furthermore, the Greek data are not the only case of mismatch emerging from work on corpora annotation. The (AmEng) ToBI (Beckman & Ayers Elam, 1997) uses Break Index 2, and the Korean and Japanese systems (Beckman & Jun, 1996 and Venditti, 1999 respectively) use the diacritic "m" to flag the presence of a phrasal boundary unaccompanied by the expected tonal events, and vice versa. The evidence emerging from the annotation of corpora corroborates previous research, such as Hyman, Katamba & Walusimbi (1987) on the intersecting domains of sandhi and tone rules in Luganda.

Data like these force us to reconsider the nature of prosodic structure and the types of events it controls and represents. Typically, prosodic structure is built on the basis of stress and/or tonal evidence, while the presence of sandhi is often considered a fortuitous coincidence. The Greek data on the other hand suggest that in a language with a wealth of sandhi rules and a simple stress system, the reverse holds: the domains of application of the sandhi rules are more complex than those of stress. This in turn suggests two possible answers: either sandhi is not really regulated by prosodic phrasing (but perhaps by morphosyntactic factors), or two distinct postlexical structures are needed, one dealing with stress (and indirectly intonation) and the other with sandhi (a solution not equivalent to using a grid and a tree respectively, since the two are not isomorphic and thus the former cannot be derived from the latter by a simple mapping).

References

- Beckman, M. & S. Jun. 1996. "K-ToBI (Korean ToBI) Labeling Convention". http://www.humnet.ucla.edu/humnet/linguistics/people/jun/sun-ah.htms

- Beckman, M. & G. Ayers Elam. 1997. "Guidelines to ToBI Labelling" (ms, OSU).
- Hyman, L., F. Katamba & L. Walusimbi. 1987. Luganda and the Strict Layer Hypothesis. "Phonology" 4: 87-108.
- Venditti, J. J. 1999. "The J_ToBI model of Japanese intonation" (ms, OSU).

'Loud' and 'soft' signing in sign language of the Netherlands: Evidence against handshape as a phonological entity

Onno Crasborn (Leiden University)

Models of sign language phonology rarely incorporate a separate phonetic implementation of the proposed features. Apparently, it is assumed that phonological feature values have obvious and invariant phonetic characteristics, thus ignoring the variation we find in sign language production, both within and between signers. The present study focuses on one type of intra-signer variation, viz. between 'loud' and 'soft' registers.

Five native signers of Sign Language of the Netherlands produced 22 signs which in their citation form have either shoulder/elbow movement or finger movement at the metacarpophalangeal (MCP, or 'base') joints. Each item was signed three times in three different situations: with an addressee present at 0.4 m. ('soft register'), 2.50 m. ('standard register'), and 20 m. ('loud register'). The data show that movements of the whole hand (path movements) can be reduced by moving the MCP joints only. Signs with MCP movement in the standard register can be enhanced by moving the shoulder, elbow, or wrist joints.

Following work on variation in speech by Lindblom (1990) and Boersma (1998), I propose that for sign language we can only explain such phonetic variation by clearly distinguishing between articulatory and perceptual representations. What the actual articulation of a perceptual phonological specification looks like is the outcome of the battle between articulatory and perceptual constraints, which varies per situation.

To account for the alternations between finger and arm movement found in this study, I propose that the phonological specification for these signs does not specify the state or movement of the MCP joints. Rather, these joints, like the arm joints, can be used to articulate path movements in certain signs in certain styles, so that the notion of "handshape" appears to play no phonological role. The more neutral term 'articulator' seems to be more appropriate (cf. Crasborn & van der Kooij 1999).

References

- Boersma, P. (1998) Functional phonology. Formalizing the interactions between articulatory and perceptual drives. The Hague: Holland Academic Graphics.
- Crasborn, O. & E. van der Kooij (1999) Base-joint position in Sign Language of the Netherlands: phonetic variation and phonological specification. To appear in: J. van de Weijer (ed.) Proceedings of HILP4.
- Lindblom, B. (1990) Explaining phonetic variation: a sketch of the H&H theory. In: W. Hardcastle & A. Marchal (eds.) Speech production and speech modelling. Dordrecht: Kluwer. Pp. 403-439.

Conflicting phonologically based and phonetically based constraints in the analysis of /1/-substitutions

Dicky Gilbers (University of Groningen)

In my presentation first language acquisition data, in particular attested realizations of /l/, are analyzed in Optimality Theory as a conflict between different types of phonologically based and phonetically based constraints. In the analysis correspondence constraints, which ensure lexical diversity, interfere with markedness constraints that prefer unmarked output structures. Phonetically based (functional) constraints that evaluate the harmonic value of e.g. perceptive and articulatory markedness are potentially conflicting mutually and potentially conflicting with phonologically based constraints that evaluate e.g. positional markedness. The different constraints reflect diverse hierarchies of markedness and systematic distances of segments and clusters. It is claimed that an adequate account of the data can be obtained neither by a purely phonologically based analysis nor by a purely functional, phonetically based, analysis. Therefore, insights from both disciplines are to be combined into one comprehensive account. Optimality Theory seems to be the perfect tool to bridge the gap between phonetic and phonological drives, since this theory enables us to formalize the various influences on the child's outputs as potentially conflicting or conspiring soft constraints on well formedness.

The data of possible realizations of the target /l/ are obtained from a case study in which the child Steven has been video-taped monthly from age 0;1 until 4;0. The corpus contains spontaneous utterances as well as data obtained from repetition tasks. The data show that /l/ is realized as [l,r,j,w,n,h] and by no other substitute segments. The ideal analysis, of course, enables us to account for exactly these substitutions and -what is equally important- no other. It will be claimed that the different realizations are caused by different influences on the child's pronunciation. In my presentation I will identify the different markedness and correspondence constraints that play a role in the possible realizations of /l/.

The markedness of /l/ can be indicated by means of several conspiring constraints based on sonority (stops and vowels are less marked than liquids) (SonMark), difficulty of articulation and lenition hierarchies. All these constraints strengthen each other in the prohibition of [l] as optimal output for /l/. If such constraints are dominant, correspondence constraints indicate the expected substitution segments for /l/.

The most interesting case is the liquid-nasal alternation as in the /l/-[n] substitution in slurf 'trunk' /slYrf/ realized as [snY{@f] (Steven (2;9)), because the data also show examples in which target /n/ is realized as [l]: snoep 'candy' /snup/ realized as [slup] (Steven (2;7)). A complicating factor is that this bidirectionality in the substitution data only occurs if the nasal or liquid is the second part of a cluster. In an onset singleton /n/ is never realized as [l].

According to Coleman (1998) and Kent et al (1996) /l/ and /n/ constitute a natural class in having spectral zero's, frequencies of sound energy loss. Although the articulation of /l/ is rather different from that of /n/, the acoustic consequences of the different kinds of bifurcation are rather similar spectra for [l] and [n]: most of the energy is in the low frequencies, with damped higher formants and frequencies of energy minima resulting from anti-formants. Therefore, in OT-terms, an acoustically based correspondence constraint will be introduced in which the difference between /l/ and /n/ is a minimal one given this minimal difference in formant positions and anti-formants. Notice that the spectral zero's are not an acoustic characteristic of the liquid /r/, which might explain the absence (or at least rareness) of /r/-[n] substitutions.

However, the bidirectionality of these substitutions is only attested in clusters, not in onset singletons. An explanation of this asymmetry has to be found in the interaction of the above-mentioned acoustically based constraint with a phonologically based constraint on the phonotactics of the language. In order to account for this asymmetry a phonological constraint Hsat is introduced, which is related to markedness constraints such as Hons and Hnuc (cf. Prince & Smolensky, 1993). Hnuc states that the best nuclear segment is an open vowel /a/, whereas the worst nuclear segment is /t/. Hons states that the best onset segment is /t/ and the worst /a/. Hsat evaluates the harmonic value of the segment in satellite position, which is the segment position between the margin and the nucleus of the syllable (cf. Cairns & Feinstein, 1982). It evaluates liquids as the most optimal satellite segments, better than nasals and glides, whereas obstruents and vowels are the worst possible candidates. The claim that liquids are better satellite segments than nasals can be underpinned on typological grounds. If a language has e.g. /kn/ clusters, it also has /kl/ clusters, but not the other way around. For example, Dutch has both /kl/ and /kn/ clusters (klop 'knock'; knop 'button'), but English has /kl/ clusters (clock) and no /kn/ clusters (knock pronounced as [nk]), at least not anymore. Hsat is the constraint that reflects this preference.

Thus, Hsat prefers /l/ to /n/ in a satellite, whereas SonMark always prefers /n/ to /l/. Consequently, the order SonMark >> Hsat enables us to describe the realization [sn¿{«f] for slurf, whereas the dominance order Hsat >> SonMark prefers [l] to [n] in satellite position ensuring the description of snoep realized as [slup]. The evaluation of Hsat is of course vacuous in all positions other than the satellite, which explains the unidirectionality of the substitutions in the margin position irrespective of the dominance order of Hsat and SonMark.

In sum, we may conclude that for an adequate analysis of /l/-/n/ alternations and their different behavior with respect to the direction of substitution in different positions, there must be an interaction of phonologically based and phonetically based constraints. Optimality Theory is an ideal theory for the analysis of this kind of data, because the theory is suited for accounts in which conspiracies of different influences are involved and the child learning its first language has to encouter various problems varying from articulatory immaturity to language-specific phonotactic restrictions.

References

- Cairns,C. & Feinstein, M. (1982) Markedness and the theory of syllable structure. Linguistic Inquiry 13, 193-226.
- Coleman,J. (1998) Cognitive reality and the phonological lexicon: a review. Journal of Neurolinguistics 11, 295-320.
- Kent,R.D., Dembowski, J. & Lass, N.J. (1996) 'The acoustic characteristics of American English' In: N.J. Lass (ed.) Principles of Experimental Phonetics. Mosby, St. Louis.
- Prince,A. & Smolensky, P. (1993) Optimality theory: constraint interaction in generative grammar. Rutgers University and University of Colorado at Boulder (ms.).

Accentual phrasing and syntactic disambiguation in Korean

Amy Schafer and Sun-Ah Jun (University of California, Los Angeles)

The effect of prosodic structure on syntactic disambiguation has received increasing attention in recent years (e.g., Nespor & Vogel, 1986; Price et al., 1991; Warren, Grabe & Nolan, 1995; Allbritton, McKoon & Ratcliff, 1996; Schafer et al., 1996; Watt & Murray, 1996; Kjelgaard & Speer, 1999). However, most of this research has been limited to studies of English. We report on a series of experiments examining the relationship between prosodic structure and syntactic disambiguation in Seoul Korean. We show that the pattern of accentual phrases, the smaller of two levels of prosodic phrases in Korean, can significantly effect the resolution of the syntactic ambiguity shown in (1).

(1)

hj@nmj@Nhan	akiM\i	appa
wise	baby-GEN	daddy

[wise baby]'s daddy N1-modification

wise [baby's daddy] N2-modification

Example (1) shows an adjective followed by a complex NP. In Korean, such adjectives can readily be interpreted as modifying either the entire complex NP ("N2-modification") or only N1 ("N1 modification"). Our items were constructed and pretested to ensure that N1 modification was always possible, but N2 modification was ultimately preferred on pragmatic grounds. Following standard assumptions in sentence processing, we predicted that listeners would initially attempt to modify N1, but would generally revise this decision when N2 was encountered. Results from a control set, which varied the pragmatic bias for N1 versus N2 modification, support this prediction. We further predicted that these interpretive preferences would be modulated by the prosodic form of the ambiguous string.

The Adj-N1-N2 string can be produced with multiple prosodies in Korean, as observed and preliminarily indicated by Jun (1994, 1996). We first conducted a production study which supported Jun's claims. Four native speakers of Seoul Korean produced sentences containing strings like (1) under conditions of wide focus and contrastive focus, as well as with the intent to disambiguate the syntactic structure. The results showed four patterns of accentual phrasing, summarized in (2). The "default" prosody placed each word in a separate accentual phrase. This contour was frequently produced for both interpretations, and was by far the most common phrasing pattern under our conditions. For intentional disambiguations, speakers tended to group the adjective with N1 for N1 modification and group the two nouns together for N2 modification. Speakers also produced all three words in a single accentual phrase, for both interpretations, when the adjective was contrastively focused.

(2)

a. (wise) (baby's) (daddy)	Default prosody
b. (wise baby's) (daddy)	N1-modification prosody
c. (wise) (baby's daddy)	N2-modification prosody
d. (wise baby's daddy)	Adjective-focus prosody

Productions from one of these speakers were then presented to naive listeners in two comprehension experiments. Experiment 1 employed a cross-modal naming task. Subjects heard an auditory fragment consisting of the beginning of a sentence through the adjective and N1 and then named aloud a visually-presented N2. Results showed a significant effect of accentual phrasing on naming times for N2. Most notably, naming times were significantly longer in the condition with N1-modification prosody than in the Default prosody condition, conditions which differ only in the presence of an accentual phrase boundary between the adjective and N1. This suggests that reanalysis to an N2-modification interpretation was particularly difficult with N1-modification prosody. It also suggests that prosodic structure has incremental effects on syntactic processing, or is retained in memory and can influence reanalysis decisions, or both: The critical prosodic difference between these conditions occurs prior to the point at which N2 is encountered, and thus prior to the point at which the parser has encountered non-prosodic evidence for reanalysis. Therefore, the Korean results lend cross-linguistic support to the argument, based on English, that a well-formed prosodic structure is built during the early stages of sentence comprehension and is available to influence subsequent syntactic and semantic processing decisions (Schafer,

1997).

Experiment 2 presented the same auditory fragments in an auditory questionnaire. Subjects answered an open-choice question about the adjective that indicated which interpretation they had constructed for the ambiguous sequence. The results show a significant effect of the presence versus absence of a prosodic boundary between the adjective and N1; N2-modification interpretations rose with the presence of this prosodic boundary. As with Experiment 1, the results suggest that the subtle phonetic cues for an accentual phrase boundary can be detected during sentence comprehension and used to help determine the ultimate interpretation of the sentence. Experiment 1 shows effects of accentual phrasing relatively early in the course of processing, using an on-line task. Experiment 2 shows that these effects persist to influence the final interpretation.

In both experiments, the results show only weak effects of the boundary between the nouns. One explanation for this is that listeners received less phonetic evidence for this boundary, since the auditory fragments ended with N1. Work in progress is investigating whether this non-effect persists when listeners receive the phonetic cues to accentual phrase structure that follow the boundary.

Auditory features for Xhosa stop types and click accompaniments

Michael Jessen (University of Stuttgart)

A representative sample of the sound system of Xhosa -- a Southern-Bantu language of the Nguni group - was investigated acoustically, with the aim of proposing a distinctive feature representation of the contrasting stop types and click accompaniments in the language.

As part of a sponsored research stay in South Africa by the author, suitable lexical items with systematically controlled tonal patterns were elicited with the help of a Xhosa informant. The word list, together with suitable contexts, was produced by four female and four male native speakers of Xhosa and recorded on DAT tape in sound-treated facilities of the host university. The data were analyzed for closure duration, voicing duration, VOT, and burst amplitude, as well as F0, F1, and voice quality (measured as H1*-H2*, etc.; cf. Sluijter, 1995; Hanson, 1997) at consecutive periods during vowel onset.

It is shown that the set of acoustic/auditory features [tense], [voice], and [checked] can capture the contrasts in a functionally explanatory, economical, phonetically motivated, and universally grounded manner. The original proposal of this feature set by Jakobson (1968) and Jakobson & Halle (1968) is supplemented with more recent ideas on and applications of these features by Kingston & Diehl (1995), Traill (1995), and Jessen (1998), together with some new interpretations of the feature [checked].

The feature [voice] is implemented in Xhosa by a wide variety of correlates, all of which are instantiations of Kingston & Diehl's "low frequency property". These include long closure voicing in the implosive and optionally (but much more rarely than in English) in the voiced stops and clicks, as well as tonal depression, low F1, and optionally a mild degree of breathy phonation (slack voice) after the voiced stops and clicks (cf. Kingston et al., 1997 on F1 and breathy/slack voice). In post-nasal context, where aspiration is ruled out, voiced stops and clicks are furthermore produced with very short oral closure durations.

The aspirated stops and clicks in Xhosa are produced not only with long VOTs but also with comparably short closure durations. This emphasizes the interaction between laryngeal and supralaryngeal components in the production of aspiration. It is argued that aspiration is more adequately represented with a feature that refers to the acoustic/auditory outcome of gestural coordination than with an articulatory feature like [spread glottis], that concentrates on a necessary but not sufficient component in the gestural makeup of an aspirated stop or click (cf. Löfqvist & Yoshioka, 1981). It is proposed that the feature [tense] serves this purpose (cf. Jessen, 1998).

Jakobson (1968) and Jakobson & Halle (1968) suggested that [+checked] is used in the representation of ejectives, implosives, and clicks. Through the inclusion of clicks, [checked] has a wider scope than the articulatory feature [constricted glottis] and therefore allows for a more economical universal feature set. It is shown how this proposal works in Xhosa and in other languages with clicks, though it was found necessary to work out some modifications of the original account of Jakobson (1968). Most sensitive to the issue are cases where ejectives have similar "click-like" auditory characteristics as clicks. In Xhosa this is the case when comparing the alveolar (abrupt) plain click with the velar ejective (provided the latter is not produced as plain k, which is possible in Xhosa). Since there is a measurable difference in the spectral energy distribution of these sounds and a perceivable tonality difference, it is argued that the feature [flat] is used to differentiate these otherwise featurally identical sounds.

References

- Hanson, H.M. 1997. Glottal characteristics of female speakers: acoustic correlates. Journal of the Acoustical Society of America 101: 466-481.
- Jakobson, R. 1968. Extrapulmonic consonants: ejectives, implosives, clicks. In Selected Writings I: Phonological Studies. The Hague: Mouton. 720-727.
- Jakobson, R. & Halle, M. 1968. Phonology in relation to phonetics. In Malmberg, B. (ed.) Manual of Phonetics. Amsterdam: North Holland. 411-449.
- Jessen, M. 1998. Phonetics and Phonology of Tense and Lax Obstruents in German. Amsterdam: Benjamins.
- Kingston, J. & Diehl, R.L. 1995. Intermediate properties in the perception of distinctive feature values. In Connell, B. & Arvaniti, A. (eds.) Phonology and Phonetic Evidence. Papers in Laboratory Phonology IV. Cambridge: Cambridge University Press. 7-27.

- Kingston, J., Macmillian, N.A., Dickey, L.W., Thorburn, R. & Bartels, C. 1997. Integrality in the perception of tongue root position and voice quality in vowels. Journal of the Acoustical Society of America 101: 1696-1709.
- Löfqvist, A. & Yoshioka, H. 1981. Interarticulator programming in obstruent production. Phonetica 38: 21-34.
- Sluijter, A.M.C. 1995. Phonetic correlates of stress and accent. The Hague: Holland Academic Graphics. Ph.D. Dissertation, University of Leiden.
- Traill, A. 1995. Place of articulation features for clicks: anomalies for universals. In Windsor Lewis, J. (ed.) Studies in General and English Phonetics. Essays in Honour of Professor F.D. O'Connor. London: Routledge. 121-129.

Listeners' representation of within-word structure: A cross-linguistic study of Japanese and English

Takashi Otake (Dokkyo University)

Investigation of listeners' segmentation of continuous speech is one of the hot issues in the phonological processing. The studies based upon reaction-time data suggest that different speakers exploit rhythm-based segmentation. The recent study attempted to reveal how Japanese and English listeners form conscious representations of within-word structure and found that both subject groups preferred syllables, suggesting that speech segmentation strategies may not be directly relevant to conscious representations of within-word structure (Otake et al., 1995).

One interesting observation in this study was that Japanese adult listeners could prefer syllables to morae, although other studies reported that Japanese listeners were sensitive to morae (Kubozono, 1995: Otake and Yamamoto, 1997). More recent studies revealed that when Japanese preschool children were asked to identify the number of chunks within Japanese words (e.g., manga), they identified two chunks (pan-da) rather than three chunks (pa-n-da) (Otake and Yoneyama, 1999: Inagaki et al., in press), suggesting that Japanese children may be sensitive to syllables. Furthermore, another recent study showed that when first graders at elementary school were assigned the same task with both Japanese and Spanish materials, their preference was given to morae in Japanese but syllables in Spanish words (Otake and Yoneyama, in press). In other words, these studies seem to suggest that Japanese speakers could be aware not only of morae but also of syllables. Although the present phonological theory assumes that syllable structures are represented in terms of syllables and morae, in which the latter is defined as a subunit of the former, few studies have been conducted to examine whether human listeners are actually aware of both units. Thus, we report four experiments with a tapping task which attempts to capture listeners' conscious perceptions of syllables and morae within-word structure.

In Experiment 1 we presented 40 2syllable English materials which contained a CVC syllable as the first syllable (e.g., panda) with the same number of filler words to 15 American college students and asked them to identify the number of chunks within each word with a tapping task. The result showed that they preferred syllables (100%) to morae (0%). In Experiment 2, we presented the same materials to the same subjects after a short training session, in which they were taught that 2 syllable words could be regarded as three chunk rather than two chunk words. After confirming that they could tap on the basis of the new instruction in the practice session, they were asked to identify the number of chunks. Yet, the result showed that they preferred syllables (75%) to morae (25%). These results suggest that English speakers may be hard to recognize a nasal in a coda as a mora. In Experiment 3 we presented 40 2 syllable Japanese materials (e.g., manga) which had the same type of structure in English with 40 fillers to 15 Japanese college students with the same task. The result showed that they preferred morae (100%) to syllables (0%). In Experiment 4, we presented the same materials to the same subjects after a short training session as we did in Experiment 2. The result showed that they preferred syllables (83%) to morae (17%).

The results in these experiments suggest that human listeners are not necessarily aware of both syllables and morae in the same manner and that the reason why Japanese listeners are aware of these two units equally well if they are asked to do so is that they may have once recognized syllables during the preschool period and shifted toward morae, while English listeners did not.

[Research supported by the Japan Society for the Promotion of Science (Grant-in-Aid for Science Research (C), #11610566) and the International Communication Foundation, Japan]

References

- Inagaki, K., Hatano, G. and Otake, T. (in press) "The effect of kana literacy acquisition on the speech segmentation unit used by Japanese young children," Journal of Experimental Child Psychology.
- Kubozono, H. (1995) "Perceptual evidence for the mora in Japanese," In Bruce Connell and Amalia Arvaniti (eds), Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV, 141-156.
- Otake, T., Davis, S. and Cutler, A. (1995) "Listeners' representations of within-word structure: A cross-linguistic and cross-dialectal investigation," Proceedings of EUROSPEECH 95, Madrid,Vol.3, 1703-1706.
- Otake, T. and Yamamoto, K. (1997) "Listeners' representations of within-word structure for bilingual Japanese speakers of English and monolingual speakers of Japanese and English," Paper presented at the

134th Meeting of the Acoustical Society of America. San Diego.
- Otake, T. and Yoneyama, K. (1999) "Listeners' representations of within-word structure:Japanese preschool children," Proceedings of the XIII International Congress of PhoneticSciences, San Francisco, 3, 2193-2196.
- Otake, T. and Yoneyama, K. (in press) "Phonological units in mental lexicon and their awareness," On'in Kenkyu 3.

The phonological status of Dutch epenthetic schwa: A challenge to articulatory phonology

Natasha Warner (Max Planck Institute for Psycholinguistics), Allard Jongman (University of Kansas), Anne Cutler (Max Planck Institute for Psycholinguistics), and Doris Mücke (University of Cologne)

In Articulatory Phonology (Browman & Goldstein, 1990, 1992, etc.), lexical representations consist of strings of gestures, and all gestures in any surface form of a word must be present in its lexical representation. Variations in surface form reflect only changes in timing or magnitude of existing gestures (although see for example McMahon et al., 1994 for a challenge to this claim). We present articulatory data showing that Dutch schwa insertion involves addition of a phonological segment, with its gestures, rather than re-timing of existing gestures.

In Dutch, epenthetic schwa can appear in clusters of a liquid and a non-coronal consonant (1). This process is variable even within speakers, and both variants are common and not sociolinguistically marked, although the form without epenthesis is more basic (Donselaar et al., 1999). Dutch also has allophonic variation between dark and light /l/ similar to that in English, and we used this allophonic variation to test the phonological status of the epenthetic schwa. Articulatory Phonology accounts for epenthesis without insertion of gestures through overlap or temporal separation of existing gestures (Browman & Goldstein, 1990; Jannedy, 1994; Gick, 1999). If gestures separate in time, a gap with no gestural specifications can appear, which would sound like a schwa (2). Such re-timing would not change the gestures themselves, so this account predicts that /l/ should have the same gestural characteristics whether an epenthetic schwa is produced or not. However, if the schwa is inserted as a phonological unit, it would form a syllable nucleus, placing the /l/ in onset position and conditioning allophonic variation in the /l/.

We used a Carstens Articulograph to record the tongue movements of 5 Dutch speakers saying words with /l/-labial clusters with and without epenthesis (3a, b) and matched words with underlying schwa (3c). Articulograph pellets were placed on the tongue tip, body, and dorsum. Speakers first read all the words naturally, then read them again deliberately with and without epenthesis. We evaluated the vertical and horizontal position of the tongue tip, body, and dorsum at the point of maximal tip raising and the point of maximal dorsum retraction during the /l/. For each speaker, 15 pairs of words produced deliberately with and without epenthesis and the matched words with underlying schwa were evaluated.

Previous research on English finds greater tip raising in light /l/ and greater dorsum (or body) retraction or lowering in dark /l/ (Sproat & Fujimura, 1993; Gick, 1999). In our data, the tip was significantly higher in /l/ before epenthetic schwa than in non-epenthesis productions for all five speakers ($p<.001$ for each speaker). The dorsum and body were significantly lower in /l/ in non-epenthesis productions than in those with epenthesis ($p<.01$ for each speaker). Tongue position in /l/ before epenthetic vs. underlying schwa was similar, and differences between these conditions were significant for only two of the speakers. Thus, /l/ is light before both epenthetic and underlying schwa but dark in words produced without epenthesis, and /l/ before epenthetic and underlying schwa is articulatorily similar. Because this comparison used productions for which subjects had deliberately produced epenthesis, we further compared productions in which subjects epenthesized spontaneously to similar words which were only produced with epenthesis under instruction (4) in order to confirm that deliberate productions were natural. Differences between /l/ tongue positions before spontaneous vs. deliberate epenthetic schwa were small and for the most part not significant.

The light/dark /l/ alternation observed in productions with and without epenthesis could not come about solely through separating the consonantal gestures in time, since there is no reason for the magnitude of the dorsal lowering gesture to decrease as it is separated from a following labial. Rather, the difference is conditioned by syllable structure, and indicates that /l/ before epenthetic schwa, like /l/ before underlying schwa, is in syllable onset position. The epenthetic schwa must therefore form a syllable nucleus, and thus must be present as a phonological segment, not simply as a perceptual epiphenomenon resulting from a period with no gestural targets. Therefore, Articulatory Phonology must allow for the insertion of gestures. We suggest that epenthesis which brings syllable structure closer to a CV pattern is in general likely to be phonological, and not a matter of gestural re-timing.

Figure 1

Figure 2

Figure 3

Figure 4

References

- Browman, C. P. & Goldstein, L. (1990) Tiers in articulatory phonology, with some implications for casual speech. In Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech (J. Kingston & M. Beckman, eds.), pp. 341-376. Cambridge: Cambridge University Press.
- Browman, C. P. & Goldstein, L. (1992) Articulatory phonology: An overview, Phonetica, 49, 155-180.
- Donselaar, W. v., Kuijpers, C., & Cutler, A. (1999) Facilitatory effects of vowel epenthesis on word processing in Dutch, Journal of Memory and Language, 41, 59-77.
- Gick, B. (1999) The articulatory basis of syllable structure: A study of English glides and liquids. Unpublished dissertation, Yale University.
- Jannedy, S. (1994) Rate effects on German unstressed syllables, OSU Working Papers in Linguistics, 44, 105-124.
- McMahon, A., Foulkes, P., & Tollfree, L. (1994) Gestural representation and lexical phonology, Phonology, 11, 277-316.
- Sproat, R. & Fujimura, O. (1993) Allophonic variation in English /l/ and its implications for phonetic implementation, Journal of Phonetics, 21, 291-311.

Sentence mode and emotional load in Bulgarian: Economy of intonational form

Bistra Andreeva and William Barry (University of the Saarland)

An earlier study [1] illuminated the role of F0 contours on the focus accent in differentiating question and sentence mode in the Sofia variety of Bulgarian.The results showed that placement of the low target of the pitch accent at the beginning and the peak at the end of the accented syllable or in the following syllable (L*+H) is critical for the perception of (syntactically and lexically unmarked) checks, which are used to confirm already known information [2]. By shifting the peak leftwards towards the beginning of the accented syllable (H*) the pragmatic category changed from check to statement. Both categories have a L% boundary tone.

However, subjects' judgements indicated that the strength of the intonational information was not equal for the two utterance types. The situational pre-context had a strong influence on the interpretation of the utterance as a check or statement. In case the intonation contour diverged from the unmarked contour for a particular function, the category judgement was accompanied by a change in the emotional message.

To examine the general validity of these observations a further experiment was carried out. Three checks, three statements with low terminal boundary tones (L%), and three statements with continuation rises (H%) were selected from Map Task recordings made for a number of male and female speakers [3]. From each of these 9 natural utterances three intonational variants were generated, one for each pragmatic category. Firstly, a stylised resynthesized version of the original (e.g. a check) was produced. Then the intonation contours for the other two pragmatic categories (e.g. statement with terminal fall and continuation rise) were derived from the stylised contour.

Four repetitions of the stimuli were presented (Roman square design) to 15 native speakers of Sofia Bulgarian in three situational contexts: question, neutral statement and polite statement. The context utterance together with the stimulus form a minimal dialogue. The natural context for the check was a statement, and for the two statement forms it was a question. In the test, each context was offered with each pragmatic category, producing potential tension between context and stimulus. The subjects were required to judge - on a five-point scale - the degree to which each stimulus was suited to its context.

The results show that all three intonational contours can be accepted as statements in the context of a preceding question, whereas the change of context cannot shift the interpretation of a statement to a check. The following explanation can be offered for this asymmetry of reinterpretation.

The context plays an extremely important role for the interpretation of checks or statements. The context priming a statement (question-answer sequence) provides enough information to uniquely specify the communicative frame. It is a strong enough speech act marker to relegate function of the intonational form to a minor one. Thus the context weakens the distinctive function that intonation has when word sequence and syntactic structure are identical.This does not, however, mean that the intonational form is irrelevant. The shift in the interpretation of the sentence mode (check to statement) can only occur because a compensatory change of modal meaning accompanies it. The check contour cannot be accepted as a neutral statement, it can only be accepted as an emphatic, impatient or angry statement. The statements with a continuation rise were also accepted in the statement context, but the compensatory modal message was of an exaggeratedly polite speaker.

Apparently, when a typical (neutral) feature of a particular communicative situation is replaced by a feature typical of another situation, it introduces an additional modal marking. Thus intonation alone, without syntactic and lexical support, can imply a certain shade of modal meaning. This phenomenon is already known at the grammatical level. In Bulgarian, for example, the future tense is the neutral form for referring to events in the future. If the present tense is used, the utterance is immediately modally marked as a firm intention. In English and German the reverse is true.

References

1. Andreeva, B. & Barry, W.J. (1997). "Intonation von checks in der Sofia-Varietät des Bulgarischen," presented at FDSL2 Potsdam (to appear in Proceedings of FDSL2).
2. Kowtko, J., Isard, S. & Doherty-Sneddon, G. (1992). "Conversational games within dialogue," Research

paper HCRC/RP-31, 1-12. University of Edinburgh.
3. Anderson, A.H., Bader, M., Bard, E.G., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J. Miller, J., Sottillo, C., Thompson, H. & Weinert, R. (1991). "The HCRC Map Task Corpus," Lang. and Speech 34(4), 351-366.

When a word is repeated--duration and reduction effects

Alan Bell and Cynthia Girand (University of Colorado, Boulder)

It is important to understand the phonology of disfluencies in order to explain phonological aspects of natural conversational speech. We examine the patterns of a specific type of disfluency: disfluent repetitions of words. We claim that, in general, when a word is repeated, both instances of the word have the same likelihood of reduction and if the pause environment is taken into account, the same durations.

We examine the durational patterns of disfluent repeated items and the influence of adjacent pauses and filled pauses from the Switchboard speech corpus. Disfluent repetitions are utterances such as

> ... to reconcile that fact which which does seem to be true ...

in which a word or sequence of words is repeated immediately and does not appear to have a function such as emphasis. Such repetitions are often associated with other sorts of disfluencies. In particular, the repetition items are frequently preceded and followed by pauses, filled pauses (uh, um), and editing terms (well, I mean, like, etc.). The results extend earlier research on disfluent repetitions to a larger, more carefully controlled body of data over a wide variety of repetition items. The data consists of 793 repetitions drawn from a phonetically transcribed portion of about four hours of conversation from the Switchboard corpus (S. Greenberg et al., ICSLP 1996) which we have carefully checked and corrected and also augmented to include portions of repetition strings that extended beyond the stretches of speech that were originally selected. Overall, there were 663 disfluent repetitions and 91 fluent ones (39 were indeterminate). The repetitions were coded for characteristics which could potentially influence durations and reductions--presence and nature of intervening items (pauses, filled pauses, editing terms), position in utterance, whether or not there is an overlapping repetition, whether the ensuing speech continued the utterance containing the repetition, and whether the last item hosted a clitic continuation (as in `it it's').

Previous work on disfluencies in general and disfluent repetitions in particular have reported effects on word durations and on the frequency of vowel reduction (Shriberg, ICPhS 1999 and references therein). It certainly appears from this work that first items are longer and less likely to be reduced than last items, although the generality and source of the differences are less clear. We also find a duration difference: the average duration for first repetition items is 258 ms while the average durations of second items is 226 ms. However, about 60% of the first items precede a pause, whereas only a bit more that 10% of the second items do. While we do not go so far as to claim a causal link between the pauses and lengthening (there may be a common source in the disfluency production), the association should be taken into account systematically.

The strongest results of the study are:

1) While first repeated items are significantly longer (258 ms versus 226 ms) than last items, there is NO independent effect once the affect of the presence or absence of following pauses is accounted for.

2) Except for item effects, there is no difference in the frequency of vowel reduction of the first and last repetition items. Moreover, the likelihood of reduction is unaffected by the presence or absence of a following pause.

3) Repetition items were longer before pauses, with the effect being somewhat stronger for pauses after last items ($p = .0002$) than after first items ($p = .005$).

4) The duration of repetition items was unaffected by the duration of following pauses.

Results 1) and 3) confirm in a more conclusive and general way earlier suggestions of Shriberg 1995 and O'Shaughnessy 1993. The second result appears to contradict the findings of Fox Tree and Clark 1997. Their study, however, was limited to repetitions of the word `the'. We, too, find significantly less reduction of first items in our `the' repetitions, so that it appears that this characteristic is not general, but limited to certain words.

We also report on some effects that are less robust (because of the smaller numbers of instances) of other characteristics of disfluent repetition strings and on some comparisons between disfluent and fluent repetitions.

On the correlates of rhythmic distinctions: The European Portuguese/Brazilian Portuguese case

Sónia Frota (University of Lisbon) and Marina Vigario (University of Minho)

In the prosodic literature on Portuguese, a rhythmic distinction is frequently noted to set apart the European (EP) and Brazilian (BP) varieties: the former has been considered a "stress-timed" language, whereas the latter has been reported to have "mixed" rhythmic patterns of the syllable and stress-timed types. By contrast, most phonetic and phonological studies show that word stress works similarly in EP and BP, and that duration differences are the main correlate of stress phenomena. The present work has two conjoint goals: (a) to contribute to a better understanding of the prosodic difference between EP and BP, namely by establishing (some of) the physical correlates of the distinction intuitively asserted in the literature; (b) to add to the understanding of the basis of rhythmic distinctions across languages.

It is well-known that the physical basis of rhythmic type differences has been difficult to ascertain (e.g. Dauer 1983). The general failure of the quest for clear acoustic or phonetic parameters that distinguish between languages with different rhythms seems to support the view that rhythmic distinctions are the result of the presence/absence of specific phonological and phonetic properties in a particular linguistic system (Dauer 1987, Nespor 1990). Among such properties, syllable structure variety and complexity as well as vowel reduction have been pinpointed as crucial. Still, this approach to rhythm does not offer an explanation for the facts of perception: speakers, adults and newborns, when exposed to filtered speech that preserves prosodic cues are able to discriminate between languages of different rhythmic types (Nazzi et al. 1998, Ramus & Mehler 1999). These findings strongly suggest that rhythmic differences are somehow encoded in the speech signal. In recent work, the examination of data from 8 languages traditionally considered to belong to 3 rhythmic classes has shown that a set of simple measures of the duration of consonantal and vocalic intervals is able to capture rhythmic distinctions (Ramus et al. forthcoming). The same method used in Ramus et al. has been applied to Portuguese materials which are comparable in size and type to those observed in that study. Importantly, EP and BP offer a particularly suitable case to examine rhythmic differences, as sentences with exactly the same words and structure can be used as materials.

Our preliminary results indicate that the two varieties of Portuguese have different rhythms: (a) the proportion of vocalic intervals in the sentence (%V) is significantly greater in BP than in EP (t-test, paired two samples, a=0,05: statistical-t=5,38, critical-t=1,73); (b) the standard deviation of consonantal intervals within a sentence (s.d. C) is significantly smaller in BP than in EP (statistical-t=4,71, critical-t=1,73). These results are as expected in the sense that, according to Ramus et al., stress-timing is correlated with a larger s.d. C and a smaller %V, whereas syllable-timing is correlated with a greater %V and a smaller s.d. C. Furthermore, these measurements can be seen as the reflection of the specific properties of EP and BP that lead to the rhythmic distinction: in EP, vowel reduction and elision result in a smaller %V and a greater variety of consonantal intervals; by contrast, in BP the absence of pre-tonic vowel reduction and the presence of epenthetic vowels contribute to a larger %V and less variable consonantal and vocalic intervals. A comparison of the BP/EP s.d. C results with those of the 8 languages studied in Ramus et al. provides additional support for the intuitive rhythmic type difference between the two varieties: sorting the languages from most to least stress-timed, EP joins the stress-timed group (English, Dutch, Polish), whereas BP is located near the opposite extreme of the scale, after Italian, Spanish, Catalan and French and close to Japanese.

Although physical correlates of the difference between EP and BP have apparently been found, at least two important questions need to be addressed: (a) which phonological and phonetic properties besides the ones mentioned may influence vocalic and consonantal intervals duration and what is their contribution to rhythm?; (b) what role does intonation play (if any) in the BP/EP rhythmic distinction? Both these questions are addressed in an extended study, currently in progress, of comparative BP/EP prosody.

References

- Dauer, R. 1983. Stress-timing and syllable-timing reanalized. Journal of Phonetics 11:51-62.
- Dauer, R. 1987. Phonetic and Phonological Components of Language Rhythm. Proceedings of the XIth International Congress of Phonetic Sciences, 268-274.
- Nazzi, T., J. Bertoncini e J. Mehler. 1998. Language discrimination by newborns: towards an understanding of the role of rhythm. Journal of Experimental Psychology: Human Perception and Performance 24(3), 756-766.

- Nespor, M. 1990. On the rhythm parameter in phonology. In Iggy Roca (ed.) Logical Issues in Language Acquisition. Dordrecht: Foris, 157-175.
- Ramus, F. e J. Mehler. 1999. Language identification with suprasegmental cues: A study based on speech resynthesis. JASA 105(1): 512-521.
- Ramus, F., et al. forthcoming. Correlates of linguistic rhythm in speech. In Cognition.

Phonology is not the only interface to phonetics: Grammatical influences on speech timing in French

Caroline Smith (University of New Mexico)

Models in which phonetics and phonology are distinct components of the grammar (e.g., Keating 1984, Pierrehumbert 1990) usually assume that the phonetics of a spoken utterance is determined by its phonological representation. That is, a deterministic phonological representation is converted to speech via a rule-governed process that decides such things as the duration of each speech sound. The phonetic part of the grammar cannot modify structure that is provided by the phonological representation, and in most such models, doesn't have access to other parts of the grammar, such as the syntax and discourse structure. Yet there is extensive evidence that speech durations are affected by a variety of factors that fall outside the domain of either phonetics or phonology. This paper focuses on some of these factors, which are clearly part of a language's grammar (ignoring other types of factors such as word frequency and sociolinguistic influences), but which do not form part of either a phonological or phonetic representation. The role of non-phonological elements of grammatical structure in determining speech durations argues for a re-evaluation of the relation between categorial linguistic structure and the gradient properties of speech. The traditional dividing line separating gradient "phonetics" from the rest of the grammar needs to be made permeable in several places, suggesting that more different components of the grammar interact than most models imply.

Previous work has identified a number of non-phonological factors that can influence the durations of speech sounds. These factors apply to various levels of structure, from an individual word to an entire discourse, and demonstrate that the entire structure of an utterance has significant consequences for the durations of the speech sounds that make it up. Some of the factors that have been found to affect durations include the lexical category of a word (Sereno & Jongman 1995); whether a word is occurring for the first time in a discourse or has been used previously (Fowler & Housum 1987; Fowler, Levy & Brown 1997); and the topic structure of a discourse around the word (Nakajima & Allen 1993).

The present study looks at a different type of grammatical information - sentence type - and shows that it, too, can affect segmental durations. This effect was tested in French, where the same string of words can be spoken as a statement or question using different intonation. Ten sets of sentences were constructed to test the effects of sentence type and intonation pattern on vowel duration. The five sentences in each set were as similar as possible, all having the same total number of syllables and ending with the same word. Within each set, two sentences were identical, differing only in their final punctuation (period or question mark); the difference in punctuation implied different intonation as well as different sentence types. Each set also included additional types of questions, that made it possible to distinguish the effect of intonation pattern (final rise or fall) from the effect of sentence type.

Six speakers of standard French were recorded reading the sentences. Comparisons were made within the sets of matched sentences of the duration and voice quality (breathiness) of the last vowel in each sentence. The results for three speakers show that the duration of the final vowel was significantly shorter for statements than for any type of question. That is, sentence type rather than intonation appeared to be responsible for the differences in vowel duration. Comparison of the breathiness/devoicing of the sentence-final vowels yielded similar findings.

These findings suggest that the grammatical factor of sentence type must be taken into consideration in a model of vowel duration, at least for French, in addition to the better-known prosodic factors such as position of the word in the sentence. This implies that such a model must be more complex than a model in which only phonological/prosodic structure and phonetic context are relevant in determining durations. The existence of these other influences, of the sort documented in this study and others, call into question the traditional notion that phonology is the sole component of the grammar to interface with phonetics.

References

- Fowler, C. and Housum, J. (1987) Talkers' signaling of 'new' and 'old' words in speech and listeners' perception and use of the distinction. J. of Memory and Language 26, 489-504.
- Fowler, C., Levy, E., & Brown, J. (1997) Reductions of spoken words in certain discourse contexts. Journal of Memory and Language 37, 24-40.
- Keating, P.A. (1984) Phonetic and phonological representation of stop consonant voicing. Language 60, 386-

319.
- Nakajima, S. & Allen, J. (1993) A study on prosody and discourse structure in cooperative dialogues. Phonetica 50, 197-210.
- Pierrehumber, J. (1990) Phonological and phonetic representation. Journal of Phonetics 18, 375-394.
- Sereno, J. & Jongman, A. (1995) Acoustic correlates of grammatical class. Language and Speech 38, 57-76.

Modeling informativeness in language processing

Michelle Gregory, William Raymond, Daniel Jurafsky, and Alan Bell (University of Colorado, Boulder)

The information value of a word has played a fundamental role in models of language production for well over a hundred years. Following Schuchardt (1885), Jespesen (1922) notes that less informative words are redundant due to their predictability. Words that are more redundant can be phonologically reduced because they are predictable to the hearer through other sources of information. More recent work has also confirmed that hearers make use of non-phonological information for word identification. Grosjean (1980) has shown that word recognition is constrained by syntactic and semantic contexts. The Cohort Model (Marslen-Wilson and Welsh 1978) accounts for these effects by allowing contextual clues to reduce the cohort size of a word, aiding in word identification. The purpose of this study is twofold: (i) We are interested in identifying and isolating some non-phonological sources of information that influence production in natural speech contexts, and (ii) we want to explore how these effects are incorporated into models of production.

(i) Numerous studies confirm earlier accounts that a word's informativeness affects its pronunciation. For example, word final t and d are more likely to be deleted in high frequency words or word pairs (Guy 1980, Bybee 1996). Jurafsky et al. 1998 and Gregory et al. 1999 have shown that the more predictable a word is from the immediately surrounding words, the more likely it is to be durationally shorter, have a reduced vowel (in monosyllabic words), or have a final t and d deleted. Larger discourse context also contributes to variation in context (Lieberman 1963, Bolinger 1981). For example, Fowler & Houssum (1987) have shown that the duration of nouns is shorter with each repetition in narratives, despite intervening referents.

These studies each assessed the contribution of only one of these factors to durational shortening. In experiment 1 we are interested in finding out if frequency, predictability from local and discourse contexts and repetition contribute independently to shortening.

The data for this study was taken from the Switchboard corpus of 120 hours of telephone conversations comprising approximately 2.4 million words. A subset of approximately 3.5 hours of speech (38,526 words) from this corpus was phonetically hand-transcribed by Greenberg et al. 1996. This transcribed speech consists of randomly excerpted utterances that can be coreferenced in the larger Switchboard corpus. Each word from this subset was tagged for rate of speech, frequency, trigram probability, duration, and a variety of other variables not pertinent to this study (Fosler-Lussier & Morgan 1999).

From this corpus we extracted 16,000 content word tokens with the contextual and probabilistic information associated with each word. We developed a multiple regression model of predictors of word duration taking into account factors known to affect duration (following pauses, filled pauses, speech rate and word length in syllable and phones). Using this regression model, we tested word frequency, word predictability, the number of repetitions in the prior context, and a measure of contextual relatedness (Landauer & Dumais 1997).

We found significant effects of frequency ($r^2 = .03\%$) and measures of conditional probabilities of the word given the preceding ($r^2=.07\%$) and following ($r^2=1.8\%$) words ($p<.0001$). Repetition of a word also decreased duration: over the range of repetitions (0-75) duration decreases by 15% ($r^2=.03\%$, $p=.003$). A higher measure of contextual relatedness also predicts a decrease in duration: over the range of semantic relatedness (0-1) duration decreases by 24% ($r^2 = .47\%$, $p < .0001$). These results show that frequency, surrounding words, and the larger discourse context all contribute independently to the predictability of words, as evidenced by their durational shortening.

(ii) Two models have been produced to explain this affect of predicability on reduction: (1) a `Speaker-Hearer model': it is `safe' for the speaker to reduce a word that is predictable because its phonological information is less important to the hearer) (Fowler & Houssum 1987); and (2) a `Purely Speaker model': the level of activation of a word in the speaker's mental lexicon directly causes predictable words to be reduced, without any model of the hearer (see Bard in press).

If a speaker is modeling hearer knowledge, then we should see durational shortening when the speaker has not previously said the word, but the hearer has. However, a two stage model predicts that while speakers try to accommodate their hearers, durational shortening is a function of activation in the production model for the hearer (Bard, in press; Dell and Brown 1991. Consequently, we should see shortening when the speaker has previously said the word, but the hearer has not.

Using only the tokens that were repeated from the earlier described dataset (n=12,638), we tested whether there was reduction due to repetition only when the word was spoken in the prior context by either the speaker or the hearer, but not both. We created a binary varialble; either the word was spoken by the speaker or the hearer. Controlling for the number of repetitions, we found that whether it was just the speaker or just the hearer that had said the word in the prior context did not matter. But interestingly, we did find that duration decreased if BOTH the speaker and hearer had already said the word. Additionally, we found that the distance to the last mention is not significant, but the repetition density is ($p<.0001$).

Because repetition by only the speaker did not significantly contribute to shortening, we conclude that speakers must be modeling their hearers to some extent. However, because previous mention by only the hearer was also not significant, we find that speakers are not considering their hearers as much as earlier findings suggest. Based on the fact that previouos mention by both the speaker and hearer was significant, we conclude that durational shortening due to repetition is not purely speaker-based or hearer-based, it is somewhere in the middle. We hope that further research of this topic will be able to tease apart speaker-based effects from hearer-based effects in a production model.

Perceived similarity and phonotactic wellformedness

Heidi Fleischhacker (University of California, Los Angeles)

This study addresses two questions about the relation between perceived similarity between underlying-surface pairings, and preference for competing surface forms.

The first question is motivated by a cross-linguistically common asymmetry in the location of epenthetic vowels with respect to word-initial consonant clusters: vowels tend to be inserted before fricative-stop (ST) clusters, but into obstruent-sonorant (OR) clusters (e.g., Broselow 1992). We hypothesized that epenthesis into OR sounds more like the non-epenthesized UR than epenthesis before OR, while the reverse is true for ST clusters. Further, we hypothesized that the fundamental factor in determining the site of epenthesis is perceptual similarity: epenthetic vowels appear where they are least obtrusive, relative to the auditory properties of the UR.

An experiment obtained similarity and preference judgments from 49 English speaking-listeners for target words and various modifications of these targets. Examples of [[target]] - [modification] pairings are in the (non-exhaustive) list below:

1. [[stok]] - [@stok] - [s@tok]
2. [[prai]] - [@prai] - [p@rai]
3. [[fl{sk]] - [fr{sk] - [f{sk]
4. [[prInt]] - [prIt] - [prIn] - [prInd]

26 similarity raters rated each modification for how much it sounded like the target word. 23 preference raters rated how much they liked each modification for the purposes of a language game in which all words must be altered.

For both similarity and preference raters, @ST > S@T and O@R > @OR (> = 'rated significantly higher than', $p<.001$). These results support the claim that asymmetrical patterns of vowel insertion - into OR, but before ST - are driven by the goal of maximal auditory similarity between underlying and surface forms.

An unexpected finding of the experiment is that similarity and preference ratings for all target-modification pairings (not just vowel epenthesis items) are closely correlated, even down to small details. This suggests that preference judgments are simply similarity judgments obtained under a different guise.

This is particularly surprising for target-modification pairings in which similarity and phonotactic wellformedness are in conflict. An example is [[prInt]] -[prInd] - [prIn] - [prIt]: [prInd] sounds more like [[prInt]] than either [prIn] or [prIt], but is phonotactically worse, having a final cluster and a final voiced obstruent that [prIn] and [prIt] both lack. Other target-modification pairings place similarity in conflict with phonotactic constraints against vowel-initial words and initial and medial consonant clusters. In each case, preference results pattern with similarity results, suggesting that preference for surface forms is determined solely by similarity to the original word, with phonotactic wellformedness playing no role.

Crucially, however, in each of these cases the phonotactic constraints involved are soundly violated in English surface forms, and do not force alternations (e.g., final devoicing, cluster simplification), although I assume, following most work on markedness (e.g., Calabrese 1995, Prince & Smolensky 1993) that phonotactic conditions of this scope are universally present in phonological systems, even if inactive. That these phonotactic constraints do not play a role in determining preference may speak more to their violable status in English, rather than supporting the conclusion that preference for surface forms is determined solely by auditory similarity to the target, with no mitigating effects of phonotactic wellformedness.

This leads to the second question raised by this study: is there really a correlation between perceived similarity between underlying and surface forms, and preference for surface forms, independent of phonotactic considerations? A second experiment is under way to address this issue, with target-modification pairings like [[{ft]] - [{Tt] - [{st], in which there is a clear conflict between similarity and absolute phonotactic wellformedness. The predicted more similar modification for target [[{ft]] is [{Tt]: [f] and [T] are highly confusable (Miller & Nicely 1955) and receive high similarity ratings (Walden & Montgomery 1975), but not so for [f] and [s]. In contrast, the only phonotactically legal choice is [{st], since final monomorphemic [Tt] clusters are impossible in English. If similarity and preference raters both rate [{Tt] higher than [{st], this would provide surprising evidence that all phonotactic considerations are outweighed by perceived similarity in determining preference for surface

forms. In contrast, if preference raters rate [{st] higher than [{Tt], this would provide evidence for a model of preference in which absolute phonotactic constraints filter out impossible surface forms, leaving a set of phonotactically possible forms, with the choice among these falling to considerations of similarity.

References

- Broselow, E. (1992). The structure of fricative-stop onsets. Ms.
- Calabrese, A. (1995). A constraint-based theory of phonological markedness and simplification procedures. LI 26: 373-463.
- Miller,G.A. & P.E. Nicely (1955). An analysis of perceptual confusions among some English consonants. JASA 27: 338-352.
- Prince, A. & P. Smolensky (1993). Optimality Theory: constraint interaction in generative grammar. Ms.
- Walden, B. & A. Montgomery (1975). Dimensions of consonant perception in normal and hearing-impaired listeners. JSHR 18: 445-455.

A phonetic rarity comes in from the cold: The phonetics and phonology of the labial flap

John Hajek (University of Melbourne), Didier Demolin (Free University of Brussels), and Ken Olson (University of Chicago)

A survey of recent major phonological texts shows a noticeable underrepresentation of the many thousands of lesser known languages spoken in Africa, Asia and the Americas. The development of phonetic and phonological theories can only be impoverished as a result. Limited access to descriptive materials, and/or native speakers has tended to hamper proper interest in these languages, and the investigation of phenomena reported in them. A clear example of such an oversight is the labial flap (transcribed here as [vb]), first reported in a Central African language as long as ago as 1907 and sporadically noted in other languages since then. Indicative of the low status and perceived rarity of the labial segment is the absence of even an IPA symbol to transcribe it. Descriptive sources provide conflicting phonetic accounts of the labial flap, although a recent study by Demolin and Teston (1996) provides important information on the phonetic characteristics of labial flaps in at least one language (Mangbetu). To date no consideration has been given to the proper phonological characterization of this segment type.

In this paper, we show quite clearly that the labial flap is not a phonetic rarity (according to Ladefoged & Everett 1997, a segment found in only 1-2 languages) but widespread: it is found in more than 60 languages across branches of a number of major language families (Chadic, Niger-Congo, Nilo-Saharan, Austronesian) in three distinct geographical areas across two continents (Africa and Asia). All the available evidence suggests that it should indeed be given full status as a phonologically contrastive segment, on a par with other segment types, e.g. uvular nasals or implosive stops, etc. Mangbetu has the following contrastive set:

/vb/
novbo 'to scratch'

/v/
novo 'metal stick'

/w/
nowo 'growing mushrooms'

/b/
nobu 'to apply mud to a house'

/b_
nOb_

/B\/
nOmB\O 'type of hornbill

/B\/
nouB\o 'hammock'

We provide new phonetic information on the labial flap across a number of languages (including Mono, Mangbetu and others), and consider the question of possible implosiveness. We also discuss in some detail its phonological characteristics, including contrastivity, allophony, phonotactic and prosodic distribution. Although no language is known to contrast more than one labial flap, variation within and across languages is possible. As a result, four types of labial flap segments have been identified so far. Particular focus is given to the feature representation of labial flaps and the implications for current phonological theory, and point to major problems, including the inability of the current feature system to capture the allophonic distribution we report.

We also include discussion of the origins and diachronic development of labial flaps, and point to evidence of what appear to be multiple pathways.

References

- Demolin, Didier & Bernard Teston 1996. Labiodental flaps in Mangbetu. Journal of the International Phonetic Association 26: 103-111.
- Ladefoged, Peter & Daniel Everett 1997. The status of phonetic rarities. Language 72: 794-800.

Initial geminates in Swiss German and Malayalam: Production and perception

Astrid Kraehenmann (University of Konstanz), Jennifer Fitzpatrick (University of Tübingen), Aditi Lahiri (University of Konstanz), and Tara Mohanan (National University of Singapore)

Initial geminates are cross-linguistically rare and theoretically problematic. With only a handful of studies in the literature [1-4,6,7], their phonetic nature remains a mystery and their phonological representation a conundrum. We report here on ongoing production and perception studies on two unrelated languages, Swiss German and Malayalam, building on our earlier phonological and historical analyses.

These languages differ in terms of the inventories and origins of initial geminates, giving several points of comparison. Swiss German contrasts singleton and geminate obstruents in initial position, e.g. [e paar] 'a bar', [e ppaar] 'a few'. Geminates are both underlying and derived through gemination across morpheme and clitic boundaries. The length distinction /p,t,k/ vs. /pp,tt,kk/ developed historically from a voicing distinction /b,d,g/ vs. /p,t,k/, respectively, which no longer exists in the language. Malayalam contrasts singleton and geminate voiceless stops, fricatives, affricates, nasals and laterals initially, e.g. [kiLi] 'bird', [kkiLi] 'tickle'. Initial geminates are all derived by word-initial vowel deletion, which is undeniably a phonological process rather than a phonetic one.

There is a large literature on geminates in non-linear phonology. Early representations of (tautomorphemic and assimilated) geminates were a single melodic unit linked to two timing slots (C or X). The subsequent moraic representation of geminates, with a single melodic segment linked to a coda mora and a syllable onset, crucially relied on the assumption that geminates are heterosyllabic. Syllable-final and syllable-initial geminates were ignored or considered phonologically unrepresentable. We disagree, and represent initial geminates in terms of quantity, with a single melody associated to two X slots, but in contrast to geminates occupying the coda, without moras (exactly like Leti initial geminates [7], and similar to Dutch long vowels [10]).

Our production studies thus far indicate that the primary acoustic distinction between singletons and geminates in all manners of articulation is closure duration, both word-medially and word-initially (see also [6,11]), as expected following numerous studies on medial geminates. The contrast is neutralized in the absence of this cue, namely for voiceless stops in phrase-initial position.

Two pilot perception experiments on Malayalam word-initial and word-medial nasals using cross-splicing and duration continua confirm that listeners use closure duration as the primary cue for consonant length. A pilot perception experiment on Swiss German phrase-initial voiceless stops confirms that in the absence of the closure duration cue, listeners are unable to distinguish length. Secondary cues such as F0, amplitude, spectral characteristics or duration of the following vowel or VOT [6,11] have as yet failed to show an influence on perception in either Swiss German or Malayalam.

In a series of perception studies on Pattani Malay, Abramson [1-4] also found that the primary cue for consonant length is closure duration. Yet unlike Swiss German, Pattani Malay listeners can perceive the length contrast phrase-initially. Abramson argues that geminate-initial disyllabic words have greater accentual prominence on the first syllable and shows that peak amplitude and relative fundamental frequency of these syllables are important secondary cues. Interestingly, neither cue is sufficient by itself, but together, both can determine consonant length in the absence of the primary cue, closure duration.

In Malayalam, while the amplitude and F0 of the vowel following a geminate is sometimes higher than that following a singleton, we have as yet found no indications that these cues are either robust or perceptually salient. Perception experiments on initial voiceless stops are currently being planned, in addition to further experiments on Swiss German stops and Malayalam fricatives and sonorants.

References

1. Abramson, A. S. 1986. The perception of word-initial consonant length. Warotamasikkhadit & Penakul (eds.) Southeast Asian Linguistics Society 4. Tempe: Arizona State University.
2. Abramson, A. S. 1987. Word-initial consonant length in Pattani Malay. ICPhS 11.
3. Abramson, A. S. 1991. Amplitude as a cue to word-initial consonant length: Pattani Malay. ICPhS 12. 98-101.

4. Abramson, A. S. 1999. Fundamental frequency as a cue to word-initial consonant length: Pattani Malay. ICPhS 14. 591-594.
5. Fitzpatrick, J. 1990. Initial geminates in Malayalam. Ms. Stanford University.
6. Fulop, S. 1994. Acoustic correlates of the fortis/lenis contrast in Swiss German plosives. Calgary Working Papers in Linguistics 16. 55-63.
7. Hume, E., Muller, J. & van Engelenhoven, A. 1997. Non-moraic geminates in Leti. Phonology 14. 371-402.
8. Kraehenmann, A. & Lahiri, A. 1999. Phonological quantity contrast in Swiss German stops: history and acoustics. Ms. University of Konstanz.
9. Lahiri, A. & Hankamer, J. 1989. Perception of length: voiceless stops in Turkish and Bengali. Journal of Phonetics 17. 283-298.
10. Lahiri, A. & Koreman, J. 1988. Syllable weight and quantity in Dutch. WCCFL 7. 217-228.
11. Local, J. & Simpson, A. P. 1999. Phonetic implementation of geminates in Malayalam nouns. ICPhS 14. 595-598.
12. Mohanan, T. 1987. Tautosyllabic geminates in Malayalam. Talk presented at the annual meeting of the LSA, San Francisco.

Implementing sound change in different Belgium French variants

Caroline Corneau (University of Brussels)

This paper wishes to identify some mechanisms in the propagation of sound change affecting a speech community. Its aim is to show how simple phonetic variations affect and are integrated in phonological systems. Its research frame is the dynamics of phonological systems. "Dynamics" refers to the changes that sound systems are submitted to. This study wants to show how sound change can be implemented in the lab, and how speakers who listen to words pronounced by speakers from another linguistic region can adopt a variant different from their own.

A starting point is to consider synchronic variation as the pre-condition for sound change (Ohala, 1989). Linked to this are the listeners' mis-parsing perceptual errors, which can be considered as a main source of phonological change (Ohala, 1981, 1989). These facts explain why particular sound changes are possible. Studying their propagation on this basis requires considering cognitive concepts in these perception-production aspects.

An important factor in the evolution of the human cognitive system and in the development of phonological systems is the mimesis capacity (Donald, 1991). Mimesis is defined as a supramodal, motor-modelling skill which creates representations that are retrievable from memory. Its function is to represent events in a conscious, intentional and deliberate way. In evolution, mimesis preceded and allowed the development of phonology and language. In a phonological context, mimesis is considered as the competence to develop and amplify variation, which shows its importance in the development of phonological systems. This study hypothesises that sound systems result from mimetic interactions between individuals - interactions that can be influenced by external factors (sociological or natural).

This study wants to examine speakers' linguistic responses in order to investigate on an experimental basis the dynamics of phenomena attributable to mimesis. Subjects perform a perception-production task. They listen to words produced by speakers from a different region than their own. Subjects repeat those words one after the other, and their production is analysed to evaluate the potential evolution of the variables. Results will be validated by acoustic measurements. A first stage was to identify variables that are likely to implement sound changes. This study examines segmental variables as factors of change resulting from amplification of decoding errors. Data is based on three regions of Belgium. Relevant variables are selected for their contrasting realisation according to the region.

Each variable is represented in 10 different words produced by speakers from one variant (A), which are repeated in random order in 20 repetition blocks. Subjects from another variant (B) are asked to repeat each word after they hear them. In a second task, the subjects are asked to imitate the words they hear. Speakers are then asked questions to establish their level of awareness for the different variables, and to see how speakers amplify variants and move away from their own. We want to know if speakers are conscious of the existing variations. This also allows us to know what clues they use to recognise and reproduce the "foreign" variant, and to differentiate between the variables that speakers think that they imitate and those that they truly reproduce. A potential development to this study is to use the subjects' production (repetition/imitation) in a perceptual judgement task, where speakers perform an identification task (i.e. find the speakers' linguistic origin) to judge the outcome of sound change produced in the lab.

First results tend to show that subjects are likely to differ in their imitating performance. Subjects can consciously reject the variant realisation of a particular variable that is submitted to normalisation and negative connotation. When the variable reflects a change in progress, subjects are likely to use their own pronunciation if it is the less conservative variant. If there is variation in the subject's pronunciation of a variable, subjects easily produce the heard variant and, in effect, amplify one of their own existing realisations of this variable, especially if it corresponds to the standard. When the variable is not consciously perceived, speakers first use their own variant, but the "foreign" pronunciation can appear in some particular items later on, and replace the first until the end of the test. Understanding these results in a cognitive perspective will require us to examine the notions of perception and awareness in the phonological domain.

References

- Donald, M., 1991. Origins of the Modern Mind: Three stages in the evolution of culture and cognition,

Harvard University Press, Cambridge.
- Ohala, J.J., 1981. The listener as a source of sound change, Papers from a Parasession on Language and Behavior, Chicago Linguistics Society. 178-203.
- Ohala, J.J., 1989. Sound change is drawn from a pool of synchronic variation, Language Change: Contributions to the study of its causes, L.E. Breivik & E.H. Jahr (eds), Mouton de Gruyer, Berlin. 173-198.

A field database for the multifaceted investigation of length contrasts

Amanda Miller-Ockhuizen (Ohio State University, Columbus)

The recognition of a vowel length contrast in the Khoisan language Ju|'hoansi are confounded by a minimality constraint that disallows monomoraic roots. Without a contrast between monomoraic and bimoraic syllables, and lacking alternations between short and long syllables, the status of monosyllabic words is difficult to deduce. I report on the collection of a database consisting of six speakers' productions of about 2000 roots. This database allowed two different types of analysis to be brought to the question of vowel length in Ju|'hoansi. The first type of evidence is distributional - looking at the statistical patterns found in the database. The patterns reveal that there are three types of roots shown in (1): (a) open monosyllables, (b) monosyllables with a coda nasal, and (c) bisyllables. Acoustic analysis of the data shows that the duration of each vowel in a bisyllabic root is about half the duration of the single vowel in monosyllabic roots. This supports my claim that roots are minimally bimoraic.

(1)

(a) Open Monosyllables

tiih	'to be heavy'
c@o	'to be wide'
!@un	'to be sick'
(kx@u	'to be dry'

(b) Monosyllables with a Coda Nasal

k@hm	'Johnson grass'
k@m	'windpump'
Zaqm	'to be thin'

(c) Bisyllables

tz@ri	'to add, increase'
n!oBe	'female monkey'

Superscripted h=breathiness on preceding vowel
Superscripted n=nasalization of preceding vowel

Earlier transcriptions of the data (Snyman, 1975; Dickens, 1995) imply that 141 out of 759 monosyllabic roots are monomoraic. In fact, these roots, while significantly shorter than those transcribed as bimoraic, have durations which overlap with the duration of those transcribed as bimoraic. There is a difference between the categorical length contrast found between bimoraic monosyllabic roots and single mora syllables of bisyllabic roots on the one hand, and the small scale difference between non-contoured and contoured bimoraic roots, where the extra length is a result of the extra duration required to realize two articulatory gestures. Distributional facts show that these 141 roots are the only roots in the language that have monotonal, monophthongal, modal vowels, and thus their transcription with a single vowel was due to the lack of necessity of having two vowels to represent the plethora of contours that are typically found on Ju|'hoansi roots.

Distributional data also reveals a variety of positional constraints, which result in each position of the word bearing separate contrasts. In onset position, labial consonants are severely limited, but the only coda consonant found in the language is a labial nasal. Likewise, voice quality distinctions are limited to the initial mora of roots, while vowel height and backness distinctions are limited to final position, with height of the initial mora being dependent on the height of the second mora. Since epiglottalized vowels cannot cooccur with high front vowels, these positional constraints allow maximal contrasts of each type to be realized. The cooccurrence constraint between high front vowels and epiglottalized vowels is grounded in the attested lowering of all formants in epiglottalized vowels compared with modal vowels. Positional constraints also allow both initial and final edges of roots to be marked prosodically.

This type of investigation represents a new methodology in field work. Carefully transcribed data sets are not enough to answer the types of questions investigated. Rather, extremely large data sets that exhibit the full set of linguistic contrasts with a large enough number of tokens of each type produced by several speakers to allow for statistical comparison of quantitative data, are necessary. Quality sound recordings must be available in order to be used in continuous acoustic analysis of the data, based on hypotheses deduced from distributional patterns. Transcriptions cannot capture different levels of contrast, such as those found in this study. As noted by

Pierrehumbert, Beckman and Ladd (Forthcoming), "the sources of categoriality cannot be understood if these tendencies are simply assumed as axiomatic in the definitions of the encapsulated models..". Investigation of such tendencies requires strictly quantitative data, which has not been encapsulated into granular phonemic transcriptions before the source of gradient differences has been fully understood.

References

- Dickens, Patrick .(1994).English - Ju|'hoan Ju|'hoan - English Dictionary. Quellen zur Khoisan-Forschung 8. Koln: Rudiger Koppe Verlag.
- Pierrehumbert, Janet, Mary E. Beckman and D. R. Ladd.(Forthcoming)."Conceptual Foundations of Phonology as a Laboratory Science". In Carr, P., Docherty, G. & Burton-Roberts, N. (eds.) Conceptual and empirical foundations of phonology Oxford: Oxford University Press.
- Snyman, J.W. (1975). Zu|'hoasi Fonologie & Woordeboek. Communication no. 37 of the University of Cape Town School of African Studies. Cape Town: A. A. Balkema.

Sound change and phonological universals in Southwest Surmic languages

Moges Yigezu (Free University of Brussels)

The Southwest Surmic languages, Narim, Tennet, Didinga, Murle and Baale are spoken in the Southern Sudan with the exception of the latter, which is also spoken in Ethiopia. Linguistically, they all belong to the Nilo-Saharan phylum in the Chari-Nile family of the East-Sudanic division (Bender, 1976).

This paper describes the phonetics and phonology of voiced geminated stops and gives a comparative perspective in order to understand and explain the evolution of geminated stops in the five languages.

Gemination is widespread both in the Surmic group and in the neighbouring Omotic and Cushitic languages of the Afro-Asiatic family. This fact suggests that it might be an areal feature. The basic descriptive facts show that in the Southwest Surmic languages almost all consonants, except implosives, can be geminated word internally. Plain stops, in particular, occur word-medially only in their geminated form. The single voiced stops also occur intervocalically but they are usually changed into voiced fricatives or approximants at this position. Sometimes they can be deleted altogether. Fortis vs. lenis labelling was used in previous literature in order to explain the complex phenomenon of gemination. In this study, however, we argue that this labelling is phonetically awkward and nearly meaningless since it describes very little of the phenomenon of gemination.

Phonetic data consists of acoustic and aerodynamic (pharyngeal pressure, oral and nasal air flows) recordings from extensive fieldwork. However, only Baale has been studied from aerodynamic perspective. Before instrumental measurements, the phonological structure of each language was carefully described.

The comparative data reveals the following facts: Intervocalically, when voiced geminated stops are changed they correspond to implosive consonants at the same place of articulation; when single voiced stops are changed their manner of articulation changes to fricatives or approximants. Sometimes the voiced stop may be deleted or prenazalized. Consider the following data:

Gloss	Tennet	Didinga	Murle	Baale	Narim
hot	ab_<ure	ab_<ure	ab_<ur	abbure	abburri
tooth	Jiggittat	NIg_<tt9t	Jig_<itaat	JIGIDaani	Nig_<ittat
tobacco	b_<ambu	b_<ab_<u	tambu	dambu	pabbu
eye	eb_<erec	xEb_<ErEc	keBerec	keerec	kEBErEc
darkness	mUGUr	mUggUr	mUUr	-------	mUwUr

These facts provide good evidence for phonological universals and lead to the following questions: Why are voiced geminated stops changed to implosives, and why the back-articulated voiced stop is more susceptible to this change? Why single voiced stops are not allowed to occur intervocalically and changed to fricatives or approximants or deleted altogether? Where did the nasal consonant in the prenazalized stops come from? All these issues are related to the Aerodynamic Voicing Constraint (AVC). It has been argued that this constraint, on maintaining voicing in stops, along with other principles can explain a multitude of phonological universal tendencies (Ohala, 1999).

The first strategy used in order to maintain voicing in stops is changing them to implosives. For aerodynamic reasons it is hard to maintain the co-occurence of velar place, the stop quality and voicing. Indeed, in order to maintain voicing during the long geminate stop closure, speakers have to actively enlarge their oral cavities that can be done by lowering of the larynx. The acoustic consequence of this gesture is to make an implosive consonant. Thus, failure to maintain voicing in stops and consequent lowering of the larynx would, therefore, result in producing an implosive consonant. Hence, we can argue that historically voiced geminated stops in Southwest Surmic languages were the source (or the origin) of implosive consonants. A similar historical development is shown by Varyani (1974) to explain the origin of Sindhi implosives.

The second strategy employed in resolving the conflict between voicing and stops is what Ohala (1983) calls unstopping the stop, that is, changing the voiced stop into a voiced fricative or even to a voiced approximant. The third strategy utilised, perhaps an extension of the second, is losing the stop altogether, i.e., deleting it, whenever maintaining voicing is no more possible because of the extinguishing of voicing. The fourth strategy, but less

frequent, resulted in prenazalization of the stop. Ohala (1999) argues that this is due to velic leakage. The air that is flowing through the glottis accumulates in the oral cavity and then venting through the velic opening during the initial part of the consonant closure thereby maintaining voicing throughout.

The problem of maintaining voicing in voiced geminated stops in Southwest Surmic languages is resolved by the four strategies discussed above but regulated by a single constraint, namely, the AVC.

References

- Bender, L. 1976 Nilo-Saharan Overview in L. M.Bender (ed), The Non-Semitic Languages of Ethiopia. Michigan State University. East Lansing, Michigan. 439-483
- Ohala, J. 1983 The Origin of Sound Patterns in Vocal Tract Constraints in F. MacNeilage (ed), The Production of Speech. Springer-Verlag, New York Heidelberg Berlin.189-216.
- Ohala, J. 1999 Phonetic and Phonological Universals. Course notes
- Varyani, P. L.1974 Sources of implosives in Sindhi. Indiana Linguistics, 35,51-54

Is vowel harmony perceptually motivated?

Khalil Iskarous, Jennifer Cole, and Hansook Choi (University of Illinois, Urbana-Champaign)

Many phonological sound patterns have been argued to have their origin in articulatory or perceptual phenomena (e.g. Linblom 1984; Ohala 1997; Flemming 1995). The phonetic optimality of these phenomena, manifested in ease of articulation or salience of perception, motivates their grammaticization. But it is not clear whether a phonetic pattern needs to be both articulatorily and perceptually favored, or whether one condition is sufficient for possible grammaticization. In this paper, we approach this problem by investigating the phonological pattern of vowel harmony. Several investigators have provided evidence that vowel harmony emerges from the articulatory phenomenon of coarticulation (Bessell 1998; Majors 1997; Przezdziecki 2000). We investigate, through 2 experiments, whether vowel harmony is also perceptually motivated. One possible perceptual motivation for vowel harmony is that consonants could be easier to perceive when they are realized as modulations on a stable vocalic background, which is what occurs, for at least one of the formants, in a harmonic domain. Another motivation for harmony could be that vowels are easier to perceive when adjacent vowels share one or more vowel-place features. We performed 2 perception experiments to test these hypotheses. In both experiments, the subjects were 17 native-speakers of American English. In the first experiment, the subjects listened to nonsense words of the form "omVCV", where the second and third vowels are one of /i,e,u,o/, and the consonant is a voiced stop from the set /b,d,g/. Each nonsense word was composed with a noise signal at two amplitude levels. The subjects listened to all combinations of V2xCxV3 as pronounced by three male native-speakers of American English at the two noise levels for a total of 282 utterances (V2(4)xC2(3)xV3(4)xSPEAKER(3)xNOISE(2)). The sounds were played at 2 second intervals and the subjects were asked to choose the consonant they heard from a list of three on a computer screen. In the second experiment, subjects heard nonsense words of the form bVCV:, where the first V is from the set /I,E,O,U/, C is from the set /t,k/, and the second vowel is from the set of long vowels /e:,o:/. The nonsense words were produced by two speakers, and composed with noise signals at two levels, for a total of 64 words (V1(4)xC(2)xV2(2)xSPEAKER(3)xNOISE(2)). The subjects were asked to identify the first vowel of each word from an orthographic list of 4 vowels.

The data in both experiments were analyzed by using a repeated measures ANOVA. For the first experiment, there were 4 within-subject variables (V2,V3,Speaker,Noise), and for the second experiment, there were also 4 (C,V2,Speaker,Noise). The dependent variable was a nonparametric sensitivity measure.

For the first experiment, our hypothesis was that if consonants are better perceived in a harmonic environment, the sensitivity measure should be high (indicating good C identification) when the flanking vowels are both front or both back. The results did not show this pattern. The mean sensitivity for V1-Front/V2-BACK was significantly higher than V1-FRONT/V2-FRONT for both "b" and "g" ($p<.05$). And for "d", sensitivity was higher when the first vowel was back, independent of whether the second vowel was front or back ($p<.05$). These results conform with those of our earlier experiment on 9 other subjects. The second experiment also had a negative result, showing that vowel identification was not significantly higher when the vowel was followed by another vowel from the same harmonic class. We found that vowel identification for the front vowels was higher when the following vowel was front, though the result was not significant. Indeed, the back vowels were also better identified when followed by a front vowel, non-significantly for "U", but significantly for "O" ($p <.05$). We are currently investigating whether there is an effect of a preceding vowel on the perception of a following vowel. Our findings offer no support for a perceptual advantage for consonsants or vowels in vowel harmony contexts. We conclude that if vowel harmony is phonetically motivated, the basis lies in the articulatory domain, and not in the perceptual domain. More generally, that it may be possible for a phonetic phenomenon to be grammaticized based only on its articulatory qualification, even when that does not lead to perceptual salience.

Gradient constraints on lexical vowel-to-vowel relations

Eleonora Calvalcante Albano (State University of Campinas, Brazil)

Inflectional languages are known to mark morphosyntactic class by means of affixes. Stems of a given class may, in turn, be identified by phonotactic constraints. Recent research (Pierrehumbert 1993, Frisch 1996) has, in addition, shown that some such constraints are best understood as gradient and directional.

This paper presents a statistical analysis of Portuguese lexical vowels to show that the verb inflection process traditionally known as 'metaphony' (Cavacas 1920) - recently renamed 'harmony' (Harris 1974) - is part of a larger, right-to-left, gradient constraint whereby stem vowels of a marked morphological class tend to agree in height. This interacts closely with another gradient, left-to-right, general constraint which increases the likelihood of stressed vowels to be low without much regard to morphology.

The data consist of the 27,074 words (4,744 of which are verb infinitives) in Ferreira (1977), an abridged dictionary of Brazilian Portuguese (BP). Both constraints probably also hold for European Portuguese (EP), since much of the relevant vocabulary is common. However, the massive reduction of unstressed vowels in EP should render their effects more opaque. The view popularized since Harris is that harmony is a morphophonological process whereby a mid vowel, penultimate to the verb stem, agrees in height with the final vowel (the conjugation "theme") when the latter is deleted as a present tense vowel suffix is added. In the second conjugation, a falling /e/ turns [E, O] into [e, o], creating alternations such as 'm[O]rte', death, 'm[o]rro', I die. In the third conjugation, a falling /i/ turns [e, o, E, O] into [i, u], creating alternations such as 'v[E]ste', clothing or s/he wears, 'v[i]sto', I wear.

Stressed mid vowels of truncated first conjugation stems may also be thought to undergo harmony (Mateus 1975), as they surface as [E, O], agreeing in height with missing /a/, such as in 'l[E]vo', I take. The data are, however, confounded by the fact that mid penults in stems exempt from "truncation" all lower under stress, regardless of conjugation, such as in 'b[E]be', s/he, it drinks (where final 'e' is the second conjugation theme).

Harris formalizes "harmony" to properly include truncation so as to rely on Kiparsky's 'Elsewhere Condition' (1973) to rationalize rule ordering. His as well as all subsequent analyses are categorical and restricted to the verb paradigm.

New light is shed on these facts if the given vowel-to-vowel relations are understood as part of general statistical trends in the Portuguese vocabulary.

"Lowering" is by far the general, unmarked case, and is best seen as a probabilistic constraint pushing the stressed vowel to be low, either absolutely or relatively to its left neighbors, throughout the vernacular. This constraint explains why [E, O] are the almost exclusive stressed mid vowels in unmarked verb forms and rarely occur in unstressed position. The "harmony" constraint signals marked, front vowel verb themes by increasing the probability - up to 1, in a truncated mid vowel stem - for preceding vowels to share the theme height.

Respecting the fact that "lowering" is a frozen, language-specific version of the universal phonetic tendency for the mandible to be lower under stress, "harmony" increases the predictability of its exceptions. Verbs with front vowel themes are not productive, but constitute a sizeable class, which includes several morphologically transparent but semantically opaque derivations, such as 'arref+ecer', to cool down, or 'imisc+uir', to meddle. The often cultivated, semantically complex, nature of these forms is indicated by their tendency to be polysyllabic and "harmonic", which contradicts the "lowering" vowel contour and trisyllabic shape of the main vocabulary.

More recently in the history of BP, the phonetics has added a new source to the lexical tendency for "harmony" to be the most tolerable exception to "lowering". A variable, phonetically gradient process tends to raise /e, o/ to [i, u] to the left of stressed /i, u/, such as in 'm[i]nino', boy, and 'b[u]nito', pretty. Though independent of morphosyntactic class, this innovation can be shown to be subject to lexical and grammatical restrictions.

Statistical and acoustic phonetic data on these seemingly old, beaten facts will have definite implications for a theory of phonological processing.

References

- Cavacas, A. A. 1920. A língua portuguesa e sua metafonia. Coimbra: University of Coimbra Press.
- Ferreira, A. B. H. 1977. Minidicionário Aurélio. Rio de Janeiro: Nova Fronteira.
- Frisch, S. 1996. Similarity and frequency in phonology. Unpublished doctoral dissertation, Northwestern University.
- Harris, J. 1974. Evidence from Portuguese for the 'Elsewhere Condition' in phonology. Linguistic Inquiry, 5(1): 61-80.
- Kiparsky, P. 1973. 'Elsewhere' in phonology. In: S. Anderson & P. Kiparsky (eds.). A Festschrift for Morris Halle. New York: Holt.
- Mateus, M. H. 1975. Aspectos da fonologia portuguesa. Lisbon: Centro de Estudos Filológicos.
- Pierrehumbert, J. 1993. Dissimilarity in the Arabic verbal roots. Proceedings of the North East Linguistics Society, 23: 367-381.

Phonological undershoot before H% in Dutch intonation contours?

Judith Haan, Linda Heijmans, Toni Rietveld, and Carlos Gussenhoven (University of Nijmegen)

Intonation contours in Dutch that end in H% come in at least four shapes: (i) one which starts low and rises at the end, L* H%, (ii) one which rises from low to mid and rises further to high at the utterance end, L*H H%, (iii) one which starts high and rises at the end, H* H%, and (iv) one which falls from high to low and rises again at the utterance end H*L H% (cf., e.g., Keijsper 1984, Gussenhoven & Rietveld 1997, Caspers 1998). In a corpus of 400 Dutch questions (200 yes-no, 200 declarative questions), almost all tokens ended in the high final boundary tone H%. In a majority of cases, this H% was preceded by the pitch accent H*L. Frequently, however, speakers used the H* H% in the same sentences, and it therefore stands to reason to assume that the L tone in H*L H% is deleted, causing the starred tone to be directly linked to H%. A similar relation may be hypothesised in the case of the (relatively infrequent) occurrences of the L*H H%, which can be interpreted as a contour which leads to L* H% by the same mechanism of PRE-H% TONE DELETION. From an articulatory point, this process could be understood as a phonologisation of undershoot, or the smoothing out of pitch movements. An implication for the morphological structure of Dutch contours might be that the process of PRE-H% TONE DELETION is semantically neutral, either because it has no effect on the interpretation or because it has the same effect when going from the unsmoothed to the smoothed contour.

A listening experiment was carried out to test the hypothesis that PRE-H% TONE DELETION has the same effect when applied to H*L H% and L*H H% in single-accent stimuli. The four contours were crossed with two preheads, high and low, to enlarge the set of conditions in which the hypothesis could be tested. Thirty-four judges evaluated the appropriateness of each of the resulting 8 contours (as presented on four sentences) on six semantic scales, QUESTION, SURPRISE, REQUEST FOR CONFIRMATION, SUGGESTION, REMINDER, and CHALLENGE. The results (cf. Fig 1) do not support an analysis whereby PRE-H% TONE DELETION has a constant semantic effect. Rather, they appear to be interpretable in terms of Ohala's (1983) universal Frequency Code, in the sense that higher pitched contours trigger meanings associated with uncertainty and submissiveness. This raises the question of whether perception experiments of the type we conducted, in which listeners judge large numbers of stimuli containing different pitch levels, is suitable for the investigation of intonational (phonological and morphological) structure, in view of the readiness with which judges resort to interpretations that are inherent in the phonetic form.

Figure 1 (left)

Figure 1 (right)

References

- Caspers, J. (1998). 'Who's Next? The Melodic Marking of Question vs. Continuation in Dutch'. *Language and Speech* 41, 375-398.
- Keijsper, P. (1984). Vorm en betekenis in Nederlandse toonhoogtecontouren. *Forum der Letteren* 25, 20-37; 113-126.
- Gussenhoven, C. and T. Rietveld (1997). Empirical evidence for the contrast between H* and L* in Dutch rising contours. In: Botinis et al (eds). *Intonation: Theory, models and applications*. Proceedings ESCA Workshop. Athens, University of Athens.
- Ohala, J. (1983). 'Cross-Language Use of Pitch: An Ethological View'. *Phonetica* 40, 1-18.

Brain responses to the processing of information structue and intonational phrasing

Kai Alter, Karsten Steinhauer, Claudia Hruska (Max Planck Institute of Cognitive Neuroscience), and Anita Steube (University of Leipzig)

The recent finding of a positive shift in the event-related brain potential (ERP) at intonational phrase boundaries suggests a specific on-line brain response to prosodic processing (cf. Steinhauer et al. 1999). In a follow-up ERP study we try to use this positive shift - called Closure Positive Shift (CPS) - as a diagnostic criterion for the absence or presence of intonational phrase boundaries in speech material containing various focus positions.

This study aims at identifying ERP correlates for the processing of both focus accents at various positions and their compatibility relative to a context question (= information structure (IS)). All sentences were preceded by context questions establishing a narrow focus on one sentence constituent.

A subset of the German sentence material (only narrow focus on the VP is presented in the abstract) consisted of 48 sentence pairs such as (A1) and (A2), where the bracketing indicates the respective intonational phrases (IPhs), capitals indicate accented syllables:

Question A1
Was verspricht Peter fuer Anna zu tun?
(WHAT does Peter promise to do for Anna?)´

compatible answer (A1):
(A1) [PEter verspricht]IPh1[anna zu entLASten]IPh2 [und das BueRO zu putzen]IPh3
(Peter promises to SUPPORT Anna and to clean the office.)

incompatible answer (A2):
(A2) [PEter verspricht ANna zu entlasten]IPh1 [und das BueRO zu putzen]IPh2
(Peter promises to support ANna and to clean the office.)

The material was recorded using the question-answer method. We will present algorithms which allow to handle the syntax-prosody mapping using different positions for a focus feature in the underlying syntactic structure.

We first applicate the notion of bracketed metrical grids and second, we associate metrically strong positions and constituent boundaries with tonal sequences, e.g., L+H* and H%/L%.

Exhaustive acoustic analysis of all 48 sentence pairs reveal that the additional IPh boundary in (A1) was prosodically realized by a pause insertion before "anna", a high boundary tone (H%) on 'verspricht' as well as by a significant prefinal lengthening of the constituents preceding IPh1. Whereas a major accent occurs on the verb "entLASten" in (A1), accentuation is shifted to the noun phrase "ANna" in (A2). These differences in accent positions could be confirmed by a locally rising pitch contour in the fundamental frequency.

The acoustic analyses in each condition confirm that the respective accent patterns (L+H*) are reflections of the IS. In addition, in (A2) the boundary seems to be prosodically 'overridden'. Note that in Steinhauer's et al. study, sentences of the type (A1/A2) show an intonational phrase boundary after the first verb 'verspricht' when they are widely focussed. In addition, acoustic data from 15 'naive' speakers show similar results for intonational phrasing.

21 subjects participated in the present experiment and judged the prosodic compatibility of questions and answers. Preliminary ERP patterns for incompatible answers (A2) show that the CPS is completely missing in (A1/A2).

To summarize, prosodic phrasing - even if it is realized in (A1) - seems not to influence the processing of IS nor it seems to be a relevant cue for the prosodic realization of IS. This fact is reflected by the absence of the CPS in (A1/A2) in the ERP data. These findings shed new light on the interaction between focussed domains and prosodic structures used for language processing.

To conclude, IPhs are less important prosodic cues during the processing of IS: The speaker uses the variation of accent positions to guide the listener's attention to the relevant parts of the message. The hearers on the other hand, use preferably accent patterns realized by pitch variation to decode IS.

Similarly, the prosodic research on focus suggests that in languages which realizes IS by means of structural variation, e.g. morpho-syntactically, focussed domains correspond to prosodic phrases (cf. Truckenbrodt 1995, Frota 1998). From a typological point of view, German seems not to take part of this language group, narrow focus realizing by means of prominence and accentuation.

References

- Steinhauer, K., Alter, K., Friederici, A.D. (1999): Brain responses indicate immediate use of prosody in natural speech processing. Nature Neuroscience, 2, 191-196.
- Truckenbrodt, H. (1995): Phonological Phrases: Their relation to syntax, focus, and prominence. PhD dissertation, MIT.
- Frota, S. (1998): Prosody and Focus in European Portuguese. PhD dissertation. University of Lisboa.

Diatopic variation of prosodic structure as a challenge for speech synthesis

Joerg Peters (University of Potsdam)

Most linguists and phoneticians will agree that further progress in speech synthesis largely depends on the successful implementation of prosodic features. Perceptual experiments have shown that accounting for prosodic features may significantly improve the naturalness of synthetic speech (e.g. Bulyko et al. 1999). Furthermore, the control of prosodic variables may enrich its potential to function as a communicative means.

In accounting for prosodic features most researchers tacitly assume that synthetic speech can be made on the model of speech data lacking diatopic variation on the suprasegmental level. The present paper shows that this assumption may not be warranted for all languages. Speech data recorded from 20 speakers born in North- and North-East-Germany were analyzed with respect to focus intonation. The results indicate that speakers from these regions use different intonational means to signal focus structure even if their speech does not show dialectal differences on the segmental level. In particular, it was found that (1) two forms of tonal reorganization are used to signal broad and narrow focus in utterances with a nuclear H*+L accent: under the narrow focus condition the speakers of one group use tone spreading whereas the speakers of the other group use tonal reassociation. Tone spreading was found to be limited to some part of the accented word which bears the focal accent. The H-tone is associated with several slots of the skeletal tier within morpho-phonologically defined units. On the other hand, tonal reassociation was found to be limited to the rhyme of the accented syllable. When the focus structure changes the H-tone will be associated with a different slot of the skeletal tier. (2) Speakers from various regions were found to differ with respect to focus intonation even when their speech did not differ by dialectal features on the segmental level. When producing utterances with broad focus, the speakers from the North were found to choose an earlier "docking-point" for the H-tone than the speakers from the North-East. When producing utterances with narrow focus, the speakers from the North were found to reassociate the H-tone whereas the speakers from the North-East use tone spreading. Thus, abandoning dialectal features on the segmental level does not necessarily imply similar changes on the suprasegmental level.

If prosodic features at least partly resist the decay of dialectal features, speech synthesis should no longer be based on the assumption that there is something like a "neutral", i.e. diatopically unmarked, form of prosodic structure for a given language. To some extent, the problem of modeling prosody in speech synthesis resembles the problem of modeling voice quality: There is no "neutral" voice quality. Rather, enhancing the naturalness of a synthetic voice requires to make it sound like the voice of an individual speaker. Analogously, we must take into account that there may be no diatopically unmarked form of the prosodic structure for a given language. Accounting for prosodic features in speech synthesis, then, may require a more flexible approach that takes diatopic variation into account from the beginning. Unlike the problem of creating naturally sounding voices, however, the present study shows that diatopic variation of focus intonation is not restricted to differences in phonetics. Both forms of tonal reorganization, i.e. tone spreading and tonal reassociation, are phonological processes. Consequently, the diatopic difference between both groups of speakers must be characterized as a difference on the phonological level.

References

- Bulyko, I.; Ostendorf, M. & Price, P. (1999). "On the relative importance of different prosodic factors for improving speech synthesis". Proc. of the 14th Int. Congr. of Phonetic Sciences, San Francisco, August 1-7, 1999, Vol I, 81-84.

Do segments of morphosyntactic words reorganise into prosodic constitutents? Evidence from English /k/ releases

Stefanie Shattuck-Hufnagel (MIT)

This study is part of an ongoing evaluation of the hypothesis that speech production planing for English involves a significant reorganization from morphosyntactic to prosodic constituents (Inkelas and Zec 1990, Hargus and Kaisse 1993, Shattuck-Hufnagel and Turk 1996). The motivation is the possibility that the morphosyntactic structure of a sentence, appropriate for message planning, is mapped into a prosodic structure more appropriate for articulation planning. In particular, can an element which is in final position in its morphosyntactic word undergo a re-mapping into initial position in a prosodic constituent, so that its affiliation shifts from leftward to rightward?

A critical question here is what acoustic-phonetic characteristics indicate a leftward vs. a rightward affiliation of a segment or syllable in a spoken utterance. Earlier observations suggest that for unambiguous cases, where the morphosyntactic and prosodic organization appear to be the same, there are reliable acoustic-phonetic differences between tokens of the same phoneme in different constituent positions, e.g. syllable-initial vs. -final, or word-initial vs. -medial (e.g. Krakow 1999, A. Cooper 1991, Randolph 1989). Such allophonic differences may also correlate with word-initial vs. word-final position of reduced-vowel syllables. For example, it has been hypothesized that word-onset consonants are produced with crisper articulation than non-onset consonants, even before reduced vowels. Thus, the word-initial /t/ of "in today's paper" may be more aspirated than the word-medial /t/ of "into the crowd". To the extent that the acoustic-phonetic shape of a segment or syllable varies systematically with its position, the variation provides evidence about the nature of the speaker's planning representation, making it of interest to determine just what kinds of structures govern this variation.

We have begun by examining the acoustic characteristics of /k/ release into a following vowel in different contexts. Our goal is to compare morphosyntactically-initial and -final tokens of /k/ in contexts where reorganization into a different prosodic structure might occur, with tokens in contexts where such reorganization is not predicted.

In the first study, we compare 1) word-final /k/ release into the monosyllabic function words "or" and "and", when the two words are located in the same intonational phrase, and therefore might be subject to cliticization of the function word leftward, with possible resyllabification of the final /k/ into initial position (Levelt 1989, Booij and Lieber 1993), and 2) word-initial /k/ release. The corpus contains utterances with /k/-final target words followed by a reduced-vowel production of "and" (e.g. "Please say Maine, or Duke and Rice will play"), and /k/-initial target words like "Canada" in similar phrases. Initial analysis of data from 3 of the 10 speakers shows that word-final /k/ is released more weakly than word-initial /k/ (i.e. the duration of aspiration is significantly shorter and its average amplitude significantly less), so that the final /k/ of words like "Duke" does not resemble an initial /k/.

These results provide no support for the possibility that the sequence "Duke and" [duk@nd] is reorganized into the syllable sequence [du] [k@nd]. However, it is likely that the release of an initial /k/ into this reduced vowel is weaker than the release into a full vowel (as in "Canada"), with or without resyllabification. Thus we plan an additional set of studies using elicitation pairs such as "dew collapsed" and "Duke elapsed", providing a better comparison for "Duke an' Rice".

In the second study we compare word-final with word-initial /k/ in utterances where the preceding or following vowel is pitch accented. This experiment explores the possibility that a word-final /k/ will resyllabify into the onset of a following pitch-accented syllable even if this does not occur for reduced vowels. Initial results for 2 of the 10 speakers suggest that it does not: the duration and average amplitude of the aspiration in both "break UP" and "BREAK up" are significantly smaller than for either "bray CUP" or "BRAY cup". This is consistent with the hypothesis that the final /k/ in "break" does not resyllabify into an initial /k/ in these utterances.

Finally, in the third study, we are examining the behavior of final-/k/ releases in several prosodically-labelled speech corpora, including the BU FMRadioNews corpus (professionally read broadcasts), the Switchboard corpus (telephone conversations between strangers on assigned topics) and the Callhome corpus (telephone conversations between family members or friends, no topic constraints). Results will indicate whether the patterns we find in laboratory speech are consistent with those in more natural conditions, in showing little evidence for resyllabification of morphosyntactic-word-final /k/ across word boundaries.

Articulatory evidence for gradient speech errors

Marianne Pouplier, Larissa Chen, Louis Goldstein (Yale University and Haskins Laboratories), and Dani Byrd (Haskins Laboratories and University of Southern California, Los Angeles)

Speech errors have long been appealed to as evidence for segmental units of encoding in speech production. It has been assumed that speech errors originate in the encoding process (not in execution) from a categorical mis-selection that shifts a segment to a wrong position within a prosodic 'frame.' Evidence supporting this notion comes from the fact that the most commonly reported type of error is single segment exchanges (Shattuck-Hufnagel & Klatt, 1979; Shattuck-Hufnagel, 1983). However, if speech errors can be shown to be potentially gradient, as opposed to always categorical, in nature, the view of segments as "swapping places" must be called into question.

Gradient errors are, superficially at least, more compatible with a view of phonological encoding in terms of articulatory gestures (Browman & Goldstein, 1992) than in terms of a segment-and-frame model. Segments are inherently categorical, while gestures have both gradient and categorical properties. They are defined by quantitative specifications of dynamical control parameters for the formation of vocal tract constrictions but also serve as discrete units of contrast by virtue of their simple presence/absence of their specification for constriction location/degree.

A severe limitation of a transcription approach to evaluating speech errors is that potential gradient errors may go unrepresented in the transcriptional record if they are obscured by other articulatory events due to the coproduction of speech gestures. In a pioneering production study of errors, Mowrey and MacKay (1990) present EMG data that suggest that gradient errors do exist, although they are not always audible. However, the anomalous activity they observe in single-motor-unit recordings does not preclude the possibility that a segmental unit was either still produced correctly via compensation by other muscle activity or omitted in its entirety in the articulatory kinematics. Using magnetometry, the present study for the first time is able to provide articulatory movement tracking in the investigation of speech errors. A subject repeated phrases like "cop top" and "sop shop" while articulatory movements of the tongue, lips, and jaw were recorded simultaneously. The experiment manipulated the rate, rhythmic pattern, and target consonants of the phrases.

Within the framework of Articulatory Phonology, which takes dynamically specified gestures to be the basic units of speech production, speech errors are predicted to exhibit two properties that distinguish them from segmental errors: (1) gradient variation in the magnitude of the "misplaced" or errorful gesture and (2) sub-segmental errors involving one but not all of the gestures comprising a segment-size unit. Errors exhibiting both of these properties are evidenced in our data. For example, during the [t] in "cop top," some tokens show a small (sometimes inaudible) raising of the tongue dorsum (which normally raises for [k] but not for [t]). This anomalous gesture varies in magnitude along a continuum of values. Secondly, "sop shop" tokens provide evidence for gestural independence. /sh/ is normally produced with two gestures, a tongue body gesture and an upper lip protrusion gesture. In errorful productions of "sop shop," sometimes both gestures constituting the segment are realized in an anomalous position. However, in other instances, the erroneous tongue body raising is not accompanied by an upper lip protrusion gesture, suggesting that the tongue body gesture participated independently in an error at a sub-segmental level.

A statistical method has been developed to characterize normal variation versus gradient errors versus categorical errors. For "cop top," for example, the distribution of the tongue-dorsum height during [k] and [t] for (error-free) non-alternating controls ("cop cop", "top top") is calculated. A gradient error during [t] of "cop top" is defined as any token whose tongue dorsum activity is intermediate between the two controls and is more than two standard deviations from the mean of both. A token that is more than two standard deviations from one control (e.g., /t/), but less than 2 standard deviations from the other (e.g., /k/) control, is classified as a categorical error.

The quantitative evaluation procedure (as well as qualitative evaluation) identifies a highly asymmetrical distribution of errors: While /k/'s constriction gesture frequently appears (erroneously) during [t], the reverse is observed only rarely. Likewise, /sh/ gestures appear more frequently during [s] than vice versa. This holds for gradient as well as categorical errors. As has been observed before in speech error corpora, this asymmetry is the opposite of what is to be expected from frequency effects, which typically play a role in speech errors (Stemberger, 1991). While Stemberger accounts for these asymmetries in terms of an "additon bias" in errors and underspecification of the encoding units, the present data are compatible with an account in terms of stable modes of inter-articulator and inter-gestural coupling.

Fieldwork in the urban jungle: An empirical phonological study of Glasgow English

James Scobbie (Queen Margaret University College, Edinburgh), Claire Timmins, Jane Stuart-Smith, Fiona Tweedie (University of Glasgow), Nigel Hewlett (Queen Margaret University College, Edinburgh), and Alice Turk (University of Edinburgh)

The rather broad phonetic transcriptions which, in the majority of cases, constitute the primary data on which phonological analyses are based ("surface representations") are increasingly recognised as inadequate sources of evidence for phonological theory. Transcription by necessity sets the original data at least one point removed from all further phonological analyses. Subsequent analysts do not have access to the speech, to the speaker, or to the circumstances of data collection. The poverty of this data has been a strong motivating factor in the development of the empirical experimental methodology for phonological research dubbed "laboratory phonology" (Pierrehumbert, Beckman and Ladd, 1996).

There are, however, still problems of idealisation even within most laboratory phonology studies. In particular, we follow Docherty et al (1995) in questioning how realistic conclusions about the phonology of an entire language or dialect can be when they are based on experimental data gathered from a small or apparently homogeneous group of speakers. One reason for an experimental methodology which eschews mixtures of speakers (of different subdialects, sociolects or idiolects) is that such data is more variable than that obtained from tightly controlled groups of subjects. Even within laboratory phonology, it would seem that there is a bias towards data which is (relatively) free of interspeaker and intraspeaker variation. Such an attitude (also common in phonetic studies as well as in traditional transcription-based work) brands variation as problematic. Yet, as we will demonstrate, variation can be a tool for investigating phonological systems, phonetic systems and their interaction.

Theoretically, within the broad church of generative phonology, a central task of is to attempt to describe the speaker/hearer's competence. But to what extent does evidence purged as much as possible of interspeaker and intraspeaker variability characterise competence? We argue that data collected without reference to some of the systematic variation that operates in the sociolinguistic domain describes a type of performance, because data collection is limited to a particular type of behaviour in a particular setting. Our view is that the limitations of native speaker consultants, descriptive grammars and experimental studies of university students or academics can be overcome through the use of sociolinguistically structured groups of subjects, because the dynamic patterns in this type of performance data make it easier to divine the underlying competence.

This paper presents results from a phonetic and phonological study of Scottish English, focussing on two well-known and uncontroversial phonological "facts" about this dialect: the existence of the extra consonants /x/ and /hw/ and the distribution of (phonetically) long and short vowels (the Scottish Vowel Length Rule). To date, relevant evidence has largely come from traditional phonological elicitation methods or experimental studies of a handful of university students. Here we re-evaluate the Scottish consonants and the SVLR, using a substantial, socially-stratified sample of speech. Wordlists containing appropriate words were digitally recorded from 32 informants, male and female, children and adults, from two areas of the Glaswegian conurbation representing broadly working and middle class backgrounds. The wordlist data were then digitised and acoustically and auditorily analysed.

Our results depend on the existence of variability in the data. We show how segmental and morphological context condition the timing of the transition between the elements of the diphthong /ai/. The segmental context also conditions the height of the first mora. Our socially stratified data reveals that the timing difference is the primary realisation of segmental environment, since vowel height itself is directly conditioned by social class membership but the timing difference is not. In the case of the consonants, which do not exist in non-Scottish varieties, we could have carefully chosen our subjects to include only those who fitted our preconceptions of dialect membership, with the consequence that we would have produced evidence to support the facts as previously stated. However, we suggest (following e.g. Docherty et al 1997) that phonological description should attempt to account for the sample as a whole. Our resulting interpretation forces a revision of the existing phonological "facts" about Scottish English. The consonants' intermittent contrastivity, their phonetic characteristics, and their low frequency of occurrence all challenge the concept of a static segmental inventory.

The analysis of such a body of evidence from a non-standard variety of English necessarily entails the description of variation, particularly sociolinguistic. We conclude by arguing that sociolinguistically balanced fieldwork has much to offer to laboratory phonology, and that our data supports the need to revise our notion of what can count

as "evidence" within phonological theory.

References

- Docherty et al. 1995. [title unknown]. Poster paper at Labphon 5, Chicago.
- Docherty, G. Foulkes P., Milroy J., Milroy L., Walshaw, D. 1997. Descriptive adequacy in phonology: a variationist perspective. Journal of Linguistics 33: 275-310.
- Pierrehumbert, J., Beckman, M., and Ladd, D.R. 1996. Laboratory Phonology. In J. Durand & B. Laks (eds) Current Trends in Phonology: Models and Methods. Vol2, 535-548.

A corpus-based analysis of French intonation

Brechtje Post (University of Cambridge)

Traditionally, French intonation has been analysed in terms of contours (Coustenoble and Armstrong 1934, Delattre 1966, 1972, Martin 1977, Léon 1993). Delattre, for instance, identifies ten different intonation contours. The distinctions are made on the basis of their grammatical function, and are reflected in differences in form. By assuming that there is a one-to-one relation between form and function, such descriptions miss some important generalisations. For instance, Delattre's question contour and major continuation contour are rises which differ in the rate at which pitch changes (i.e. a concave as opposed to a convex shape of the rising contour). However, experimental evidence has shown that this difference is not systematic (Di Cristo 1976, Rossi 1978). Also, some linguistically relevant differences cannot be described, because the contours are essentially holistic units, which are only vaguely associated with the text. Thus, consistencies in the timing of the turning points of the contour cannot be expressed (e.g. a rising-falling contour with a peak on the penultimate syllable cannot be distinguished from one in which the peak is located on the final syllable of the utterance).

More recent descriptions remedy these shortcoming by analysing French intonation as sequences of High and Low tones which are connected through phonetic transitions to form the surface melodic contour (Hirst and Di Cristo 1984, Mertens 1987, Post 1993, Jun and Fougeron 1995, Di Cristo and Hirst 1996, Jun and Fougeron to appear). Since the tones are associated with specific syllables (stressed syllables, domain boundaries), the location of the turning points in the contours can be referred to, and timing differences can be captures. Also, the similarity between intonational forms, such as Delattre's question and major continuation rises, can be expressed more transparently, because the contours are decomposed into separate tonal elements. As a consequence, fewer phonological categories are needed to account for the intonational forms.

In Mertens' account, every syllable is associated with a tone, which is uneconomical. That is, if only accented syllables and/or domain boundaries are specified for tone, the unaccented syllables can automatically be assigned their correct values in the course of phonetic implementation (cf. Pierrehumbert 1980). Jun and Fougeron take an approach that is very similar to Pierrehumbert's (1980). According to Jun and Fougeron, French intonation is organised into three prosodic levels: the Accentual Phrase (AP), the Intermediate Phrase (ip) and the Intonation Phrase (IP). Each domain has its own tonal specification: LHiLH* for the AP, L- or H- for the ip, and L% or H% for the IP. These tonal specifications are aligned with the text and interpreted in terms of fundamental frequency at the phonetic level of representation. Not all phonological tones are realised; they can be phonetically undershot or transformed. LHiLH*, for instance, can also surface as LLH*, HiLH* LHiH*, LH* and LHi(L)L*. However, since these forms only arise in the phonetics, they must represent allophonic variants. As a consequence, *c'est fantastique* with a pitch peak on *fan-* and another on *-tique* must be phonologically identical to a realisation with only one peak on *-tique*, a prediction which does not seem to be correct. Hirst and Di Cristo's approach is closer to that taken by Gussenhoven (1984) for English in that they distinguish between an underlying and a surface phonological representation. The underlying tones are transformed by means of a set of phonological rules. However, although most of the discretely different contours only arise after the application of some of the rules, other rules in fact never give rise to phonological distinctions. Thus, the account blurs the distinction between the phonetics and the phonology, and its predictions are not always clear.

In this paper, I will present an autosegmental-metrical account based on Post (1993) in which two levels of representation - phonological and phonetic - suffice to capture the French data. The following tonal primitives are specified in the phonology: (a) the pitch accents H* and H+H*, (b) the boundary specifications L%, H% and 0% (i.e., unspecified for tone), and (c) an L-tone, which is optionally inserted between two high starred tones. These elements combine into phonologically contrastive intonation contours when they are associated with the segmental structure. The grammar in (1) gives the tonal specification of an Intonation Phrase. The tonal string is interpreted in terms of fundamental frequency and time alignment of the phonetic targets. The targets of H* tones that immediately follow a high tone (including H+) are automatically lowered (i.e. !H*), unless they are followed by a high boundary tone. Spreading accounts for phonetic variants in which pitch continues at the level of the preceding tonal specification.

The account is based on production and perception data (Post to appear). Two carefully constructed corpora of read and spontaneous speech were analysed auditorily and acoustically to identify the phonological contrasts. First, the pitch movements were compared across speakers and contexts, which resulted in a classification into discretely different contours. Next, the contours were compared with those described in the literature, and an inventory of

contrasting contours was drawn up. Figure 1 gives the contrasts that were observed in the speech corpora. The stylised contours abstract away from realisational differences in, for instance, the direction of pitch on the preaccentual syllables (onset) and the timing of peaks and dips.

I will argue that the present account is more parsimonious and transparent than competing autosegmental accounts of French, and that it allows us to make clear predictions about discreteness in variations in pitch, which can be experimentally verified.

---

(1) The Intonation Phrase:

%L	$(H^* (L))_0$	H*	L%
%H		H+H*	H%
			0%

(curly brackets contain the set of tones available in the given position, parentheses indicate optional elements, and H*(L) can be repeated on any non-final stressed syllable)

Figure 1

References

- Coustenoble, H. and L. Armstrong (1934). *Studies in French intonation*. Cambridge: Heffner.
- Delattre, P. (1966). Les dix intonations de base du français. *French review* 40 (1), 1-14.
- Delattre, P. (1972) The distinctive function of intonation. In: Bolinger, D. (ed.), *Intonation*. Harmondsworth: Penguin Books, 159-174.
- Hirst, D. and Di Cristo, A. (1984). French intonation: A parametric approach. *Die Neueren Sprachen* 83 (5), 554-569.
- Di Cristo, A. (1976). Indices prosodiques et structure constituante. *Cahiers de Linguistique, d'Orientalisme et de Slavistique* 7, 27-40.
- Di Cristo, A. and Hirst, D. (1996). Vers une typologie des unités intonatives du français. *Proceedings XXI J.E.P, Avignon 1996*, 219-222.
- Jun, S.-A. and Fougeron, C. (1995). The accentual phrase and the prosodic structure of French. *Proceedings ICPhS* 13 (2), 722-725.
- Jun, S.-A. and Fougeron, C. (to appear). A phonological model of French intonation. In: Botinis, A. (ed.), *Intonation: Models, analysis and applications*. Cambridge: Cambridge University Press.
- Gussenhoven, C. (1984). *On the grammar and semantics of sentence accents*. Dordrecht: Foris.
- Léon, P. (1993). *Phonétisme et prononciations du Français*. Paris: Nathan.
- Martin, P. (1977). Résumé d'une théorie de l'intonation. *Bulletin de l'Institut Phonétique de Grenoble VI*, 57-87.
- Mertens, P. (1987). *L'intonation du français. De la description linguistique à la reconnaissance automatique*. Doctoral dissertation, KU Leuven.
- Pierrehumbert, J. (1980). *The phonetics and phonology of English intonation*. Doctoral dissertation, MIT.

- Post, B. (1993). *A phonological analysis of French intonation*. Unpublished Master Thesis, University of Nijmegen.
- Post, B. (to appear). *Tonal and phrasal structures in French intonation*. Doctoral dissertation, University of Nijmegen.
- Rossi, M. (1978). La perception des glissandos descendants dans les contours prosodiques. *Phonetica* 35, 11-40.

A prosodic algorithm for text-to-speech synthesis

Karijn Helsloot and Barbertje Streefkerk (University of Amsterdam)

It is well-known that phrasal prosody is one the most difficult issues in making rule-synthesized speech as naturally sounding as normal speech. The synthesizer must be instructed to locate and quantify properly main word stress, secondary word stress, focus stress, intonation contours and intonation boundaries. Phrasal phonology (cf. Selkirk 1984, Beckman & Pierrehumbert 1986, Inkelas & Zec 1990, Hayes & Lahiri 1991, Helsloot 1995, Ladd 1996) flourished abundantly in the past two decades, but speech technology profited poorly from these theoretical developments. In this talk we present an algorithm for text-to-speech synthesis translating relevant morpho-syntactic information, lexico-semantic information and purely rhythmic principles into a prosodic framework. The results give rise to a metrical grid representation with a maximum of five levels which, in turn, must be translated in terms of the acoustic features Fo, intensity and duration .

The analysis so far is based on a corpus of 1244 Dutch read sentences, coming from the Polyphone Corpus (KPN/SPEX). Patterns of prominence are obtained by judgements of ten naïve listeners. These judgements lead to a prominence scale from 0 to 10: a word has 0 prominence when none of the listeners perceive the word as bearing prominence, and a word has 10 prominence when all listeners perceive the word as bearing prominence, and so on (see Helsloot & Streefkerk 1998, Streefkerk et al., 1998). The algorithm that we developed contains word-related components and sentence-related components. The former refer to word classifications and to mono- versus polysyllabicity, the latter to the beginnings and endings of sentences, to prosodic headedness in modifier-head NPs, and to the eurhythmic principles no clash and no lapse. The algorithm is summarized in (1).

(1)

(1a) word-related components	**(1b) sentence-related components**
1. x to each syllable	5. sentence-initial trochee
2. x to each polysyllabic word	6. sentence-final trochee
3. x to {N, V, A, Adv, Num, Neg}	7. trochee on {A-N}
4. x to {N, A, Num, Neg}	8. no stress clash
	9. no stress lapse

Application of the algorithm incrementally builds a metrical grid representation. More precisely, the word-related components build a grid representation, and the sentence-related components readjust this grid representation. An example is given in (2), in which application of some of the components is illustrated:

(2)

5.		x					
3.+4.		x			x		
2.+3.		x		x		x	x
1.	x	x		x	x	x	x
	ik	luisterde		hoe	de	wind	blies
prominences	0	8		0	0	7	2

The table in (3) gives the statistical results of the implementation of the algorithm for the analyzed corpus. For each metrical grid level the mean value of perceived prominence and its standard deviation are calculated. A completely stressless syllable (1-x) has a mean prominence value of 0.33, the syllable most heavily stressed (5-x) has a mean prominence value of 7.38.

(3)

Perceived Prominences

	Mean	Standard Deviation
1-x	0.33	1.21
2-x	2.09	3.02
3-x	3.97	3.20

4-x	5.58	2.80
5-x	7.38	2.31

In our presentation we also hope to discuss the results of the application of the algorithm to a corpus of spontaneous Dutch speech.

References

- Beckman, M. & J. Pierrehumbert 1986. "Intonational structure in Japanese and English." Phonology 3:255-309.
- Hayes, B. & A. Lahiri 1991. "Bengali intonational phonology." Natural Language and Linguistic Theory 9:47-96.
- Helsloot, C.J. 1995. Metrical Prosody. HIL dissertation 16. Den Haag, HAG.
- Helsloot, C.J. & B. Streefkerk 1998. "Perceived prominence and the metrical-prosodic structure of Dutch sentences." IFA Proceedings 22: 95-109.
- Inkelas, S. & D. Zec (eds.) 1990. The phonology-syntax connection. Chicago, UCP.
- Ladd, D.R. 1996. Intonational phonology. Cambridge, CUP.
- Selkirk, E. 1984. Phonology and syntax. Cambridge, Mass., MIT Press.
- Streefkerk, B. M., Pols, L. C. W. and Ten Bosch, L. F. M. (1998) "Automatic detection of Prominence (as defined by listeners' judgments) in read aloud Dutch sentences" Proc. of ICSLP-98, Vol 3, 683-686, Sydney.

Experience vs. psychoacoustics in the perception of vowels contrasting for [ATR]

John Kingston (University of Massachusetts), Didier Demolin, and Alain Soquet (Free University of Brussels)

Vowels which contrast for tongue root advancement [ATR] often differ in voice quality, too: [+ATR] vowels are pronounced with a lax or breathy voice quality while [-ATR] vowels are pronounced with a tense or creaky voice quality (Denning, 1989). When voice quality and tongue root position covary in this way, they affect the vowel's low-frequency center of gravity (COG) similarly. As voice quality varies from tense to lax, energy falls off faster with increasing frequency in the source spectrum, and as the tongue root varies from retracted to advanced (or as the pharynx varies from contracted to expanded), F1 gets lower. Therefore, COG is very low in lax voiced, advanced tongue root vowels (L-A Vs) vs very high in tense voiced, retracted tongue root vowels (T-R Vs). Psychoacoustically, this covariation of source and filter articulations could thus produce particularly "dull" vs "bright" vowels. Kingston, et al. (1997) showed that American English listeners discriminate synthetic dull L-A from bright T-R vowels far better than half-dull and half-bright L-R and T-A vowels. Speakers of languages in which vowels contrast for [ATR] may therefore combine a lax or breathy voice quality with an advanced tongue root and vice versa because the resulting difference in dullness vs brightness is psychoacoustically salient. However, the articulation and acoustics of the tense:lax contrast between American English vowels resembles the [ATR] contrast: The tongue root is more advanced and F1 is lower in tense than lax vowels (Perkell, 1969; Baer, et al. 1988; Jackson, 1988), and tense vowels are pronounced with laxer voice qualities than their lax counterparts (Bloedel, 1994). The same dullness:brightness dimension may therefore be as perceptually important for the perception of the [tense] as the [ATR] contrast.

To determine whether the perceptual interaction between the acoustic correlates of voice quality and tongue root position depends on linguistic experience vs the psychoacoustic salience of dullness vs brightness, we are examining how well French speaking listeners discriminate vowels in which these two articulations vary orthogonally. French vowels contrast for neither [ATR] nor [tense], so these listeners have no linguistic experience that would make one pair of vowels more discriminable than any other. Only the psychoacoustic salience of the resulting dullness:brightness differences could make the difference between L-A and T-R vowels more discriminable than the opposite combination of source and filter articulations for these French listeners. And therefore the most parsimonious explanation for their covariation in languages whose vowels contrast for [ATR] or [tense] would be that speakers of these languages combine those source and filter articulations which produce the greatest differences along the dullness:brightness dimension.

To demonstrate that the perceptual interaction always depends on psychoacoustics rather than linguistic experience, regardless of whether the listeners' language contrasts vowels for [ATR], [tense], or neither feature, we are presenting the same stimuli to listeners who speak Luo or American English as well as to French listeners. The stimuli are of the form [VkV], in which the first vowel is either a [+ATR] or [-ATR] mid front vowel, /e/ or /E/, and the second vowel varies between [+ATR] and [-ATR] high front or back vowels, /i-I/ or /u-U/. Formant frequencies and bandwidths are derived from Jacobsen's (1978) Luo data and Demolin's (unpublished) Mangbetu data; voice quality differences are synthesized by manipulating open quotient and spectral tilt in the way Kingston, et al. (1997) did. The first vowel serves as the context for judging the second vowel. The context vowel is mid because mid vowels contrast for [ATR] in all languages in which this contrast occurs, and this contrast's acoustic correlates generally differ most in mid vowels. Voice quality covaries with tongue root position in the context vowels just as it does when these vowels are produced naturally, and they are thus maximally dull or bright by design. On the other hand, voice quality varies orthogonally with tongue root position in the target vowels.

Listeners perform two perceptual tasks: discrimination of the target vowels and judgment of the difference between the target and context vowels. If listeners from all three languages find dullness-brightness differences equally salient, then they will all discriminate best those stimuli in which a lax voice quality combines with an advanced tongue root and vice versa. The context vowel is not expected to affect discrimination of the target vowels by the American English or French listeners, but it could make discriminating the target vowels harder overall for the Luo listeners who expect the target vowel to have the same [ATR] value as the context vowel because vowels harmonize for [ATR] in Luo words. Nonetheless, Luo listeners should still find vowels differing in dullness-brightness more discriminable. Judging the difference between the target and context vowels can't be done on the basis of filter characteristics alone because they always differ in height and for /u-U/ in backness as well. The question is whether the difference judgment is instead based on comparing the dullness-brightness of the two vowels. Luo listeners are most likely to use this dimension because that is precisely the phonetic dimension along which vowels harmonize in that language. American English listeners may do so, too, because voice quality

covaries with tongue root position in their vowels. However, French provides no phonological incentive to judge vowels with respect to this dimension. Linguistic experience thus predicts that listeners from the three languages should differ in how well their judgment of the difference between target and context vowels can be predicted from dullness:brightness differences between them, while this difference's psychoacoustic salience predicts instead that listeners from all three languages will use it equally.

References

- Baer, T., Alfonso, P. J., & Honda, K. (1988). "Electromyography of tongue muscles during vowels in /@pVp/ environment," Ann. Bull. Res. Inst. Logoped. Phoniat., U. Tokyo, 37, 7- 18.
- Bloedel, S. L. (1994). An Analysis of the Acoustic Correlates of Breathy Phonation in the Speech of Adult Men and Women and Pre-pubescent Males, M.S. thesis, U. Wisconsin, Madison.
- Denning, K. (1989). The Diachronic Development of Phonological Voice Quality, Ph.d. Dissertation, Stanford U.
- Jackson, M. T.-T. (1988). Phonetic Theory and Cross-linguistic Variation in Vowel Production, Ph.D. dissertation, UCLA (also UCLA WPP 71).
- Jacobsen, L. C. (1978). DhoLuo Vowel Harmony: A Phonetic Investigation, Ph.D. dissertation, UCLA (also UCLA WPP 43).
- Kingston, J., Macmillan, N. A., Walsh Dickey, L. Thorburn, R., & Bartels, C. (1997). "Integrality in the perception of tongue root position and voice quality in vowels," J. Acoust. Soc. Am., 101, 1696-1709.
- Perkell, J. S. (1969). Physiology of Speech Production: Results and Implications of A Quantitative Cineradiographic Study, Cambridge, MA: MIT Press.

A prosodic study on focus perception

Kyoko Nagao (Konan University, Kobe)

When people put emphasis on certain words, they may speak more loudly, or speak more slowly, but in most cases, it is said that speakers raise their pitch. It has been supported by numerous studies that the fundamental frequencies (hereafter f0) increase in the emphasized words, whereas decrease in the rest of speech (Pierrehumber & Beckman 1988, Kori 1989). When the f0 differences produced by a speaker become prominent enough to the listener, listeners must be able to interpret what the speaker meant. However, the observable differences in speech cannot always be perceived by listeners as they are expected. It has not been well examined how accurate a listener can interpret the words as emphasized. Furthermore, we do not know how many amounts of f0 differences are required to become noticeable to listeners. The quantity of the acoustic differences depends on various factors of the speakers, such as their morphology, and their sociocultural or psychological backgrounds. In Japanese, for example, when speakers put focus on a certain word, females showed smaller increase in the proportional f0 values on the focused words than males did, when viewed in proportion to the speakers' normal f0 range (Nagao 1999). Smaller increase seen in female speakers of Japanese are explained by gender-related constraints on (higher) pitch, which is psychologically associated with femininity in many cultures. Because of stronger sociocultural expectations for women in Japan, Japanese female tend to employ higher frequencies within their pitch range. Therefore, it can be thought that there is less room for focus expansion for female speakers. Since the non-linguistic factors such as gender have been left in the other fields of study, linguistic theories may have overgeneralized the quantitative differences caused by speakers' physiological differences. The main purpose of this study is to examine whether listeners are able to interpret speakers' intention.

The perceptual experiments were conducted using Japanese. Each utterance was spoken in Japanese by both males and females with contrastive focus on a particular word. The test words included unaccented and accented words. The listeners in one group were asked to listen to a sentence and to select the correct interpretation from the alternatives. The listeners in another group were asked to listen to a pair of sentences and select the more emphasized one. Each pair has a sentence with focus on one word and the same sentence without focus.

The listeners can select the focused one when the focused word was accented, whereas listeners showed poor performance for the unaccented words. There could not be seen significant differences between male and female speakers. As for the results from the test interpreting sentences, the listeners did not show good performance as seen in the detection test. Although there are salient f0 differences in speech, the listeners could not interpret what the speakers tried to tell them as precisely as we expected. When temporal differences such as pause before the focused word or lengthening the focused word were added to the f0 differences, better performances were observed. The results suggest that the primary cue in speech production does not always play the same role in speech perception.

References

- Kori, S. (1989). Kyocho to intonation. In M. Sugito (ed.), Koza Nihongo to Nihongokyoiku 2, Meiji Shoin, Tokyo, 316-342.
- Nagao, K. (1999). Gender differences in fundamental frequency in focused words: A case from Japanese. In Proceedings of the 16th International Congress on Acoustics and the 138th Meeting of the Acoustical Society of America, Woodbury, NY: Acoustical Society of America.
- Pierrehumber, J. B., and M.E.Beckman. (1988). Japanese Tone Structure. MIT Press, Cambridge, Massachusetts.

Phonological encoding in a whistled language: A study of Silbo Gomero

Annie Rialland (CNRS, Paris)

This paper presents a phonological, acoustic, and perceptual study of Silbo Gomero, a whistled language of La Gomera, one of the smaller Canary Islands. Silbo Gomero is a whistled encoding of Spanish as spoken on the island. Until about twenty years ago, Silbo Gomero was used as a mean of long-distance communication by all the inhabitants of the mountainous interior to convey a large variety of messages (indeed any message at all, according to La Gomera residents). There are only about 50 fluent whistlers of Silbo Gomero today, though it is undergoing a small revival and has recently been introduced into the curriculum of some primary schools.

The present study is based on data collected on La Gomera, consisting of recordings and films of whistled exchanges, interviews, and perception tests involving four whistlers. A part of the data has been presented in a televised documentary [1].

This paper is organized in three parts: 1) an overview of phonological encoding in Silbo Gomero, 2) a comparison of the whistled "melody" in Silbo Gomero utterances to speech formants (F2 and F3) of the encoded Spanish message, and 3) remarks on Silbo Gomero as a real-life perceptual experiment.

1. The whistled signal is much poorer than the speech signal, as it consists mainly in a single sinusoid whose frequency fluctuates between 1000 Hz and 3000 Hz, with ou without some harmonics. The challenge is to encode sequences of Spanish phonemes into this signal with minimal loss of distinctivity.

We first show how and to what extent this whistled language responds to this challenge. We will show how Spanish phonemes are realized in various contexts, using our own data as well as that of our predecessors (Clase 1976, Trujillo 1978). It will be shown, on the basis of production and perception tests and interviews, that some distinctions are easy to realize, some more difficult, and some impossible (leading to a small number of phonemic mergers).

Silbo Gomero is based upon acoustic cues extracted from speech. The main acoustic features of spoken Spanish encoded in Silbo are the following :

- F2 (and to some extent F3) trajectories, which form the basis of the melodic contour
- the silence of voiceless stops
- the drop in intensity during the closure of voiced consonants
- the friction noise of fricatives
- the brief closures of flaps and trills

Silbo Gomero combines these cues in its encoding of the Spanish phonemes. In addition, certain features of intonation are present in some utterance types.

2. Though F2 contours of Spanish utterances are transmitted in Silbo Gomero, their values in Hz are not precisely reproduced, as is shown by the following comparison of the F2 and F3 values of spoken vowels with the values of their whistled counterparts:

spoken vowels:

	i	e	a	o	u
F2	2325	2050	1410	920	610
F3	2950	2615	2100		

whistled vowels:

2700 2220 1550 1400 1400

Generally speaking, the values of whistled vowels are higher than the F2 values of spoken vowels. Furthermore, the Spanish vowels /o/ and /u/ are not distinguished in the whistled language. In the case of the unrounded vowels /i e a/, these differences can be explained on the assumption that it is not F2 but F2' (a perceptual equivalent of the set of F2, F3, and F4, cf. Stevens 1998) that is being reproduced. A further factor in the case of the rounded vowels /o u/ is the constraints imposed by the floor of the pitch range employed by each whistler (around 1200 Hz for the whistler illustrated above), which do not allow the reproduction of very low F2 values; this constraint explains the

merger of /o/ and /u/.

3. Silbo Gomero as a "real-life" laboratory experiment. Many experiments on "duplex" (i.e. linguistic vs. non linguistic) perception use formant like-pure tone glides mimicking the patterns of formant trajectories (Bailey 1999, Best and al. 1981, Liberman 1995). Silbo Gomero shows that the simplest stimuli (single formant-like glides) can be perceived in a linguistic way, provided the listener has a knowledge of the whistled language.

Silbo Gomero also provides an experimental test of the conclusions of numerous experiments, such as those performed with pattern playback (Delattre and al., 1954), that F2 trajectories are sufficient for the identification of consonantal place of articulation. Silbo Gomero suggests that F2' (a transform of F2, F3, and F4) may provide a better cue in real-life speech recognition tasks in which messages must be transmitted over long distances against background noise.

4. Conclusion. The paper concludes that the code of the whistled language is basically identical to that of Spanish, except for the severe reduction of the acoustic information transmitted.

References

1. Jampolski M., "Les derniers siffleurs de la Gomera", broadcast on the Arte cultural channel in France, November 11, 1999.

Positional Effect in Phonetic Grammar

Eon-Suk Ko (University of Pennsylvania, Philadelphia)

Many languages exhibit asymmetries with respect to contexts in which phonetic contrasts can be realized. Recent studies of Optimality Theory (OT) have implemented such asymmetries in the notion of positional faithfulness (Beckman 1997).

In this study I show that such a positional effect is in action not only at the level of phonology but also in the phonetic module of the grammar (Keating 1985, Cohn 1990, and Hubbard 1998). Evidence for this argument is based on a phonetic phenomenon of Chonnam Korean that has been wrongly interpreted as phonological, i.e. vowel lengthening in the word initial stressed syllable.

I contend that although this phonetic effect is sensitive to a certain position, it cannot be explained as a faithfulness effect since its surface form cannot stand in a correspondence relation with the UR, but can access only the output of phonology due to its gradient nature. Therefore, this study provides support for the existence of a level which coincides with the output of phonology and the input to phonetic interpretation, entailing a conclusion for a strict separation of phonetic and phonology.

In Chonnam, there are minimal pairs distinguished by the presence/absence of a lexical accent on the initial syllable, which is realized with increased vowel duration and amplitude on surface. When there is no lexical accent on the initial syllable, stress is realized on the second syllable by default, but no such effect is found on a non-initial stressed syllable.

The longer duration of a vowel as a result of the prominence in initial syllable has been traditionally analyzed as a phonological vowel length that is bimoraic. However, the following experiment shows that the long vowel in Chonnam is not a bimoraic segment but is an acoustic correlate of stress in a syllable-timed language.

In an experiment on the relation of phonetic and phonological timing, Chonnam was compared with Modern Seoul Korean and Tokyo Japanese. Seoul was chosen since it is known to have no vowel length distinction (Magen and Blumstein 1993), and Tokyo was chosen since it is one of the prototypical moraic languages. Minimal pair data traditionally known to be distinguished by vowel length were collected from 3 subjects for each language. For Chonnam and Seoul, subjects were asked to read 10 minimal pairs three times, and for Tokyo, 22 minimal pairs twice. For all data, the target vowel was in the initial syllable.

The results are summarized as follows: (1) The ratio of the vowel duration between the short and long series was on average 1.3 (Chonnam), 0.99 (Seoul), and 2.7 (Tokyo). (2) The ratio of the overall syllable duration was on average 1.0 (Chonnam and Seoul), and 1.6 (Tokyo). (3) The ratio of the amplitude of the vowel was 1.4 (Chonnam), 0.98 (Seoul) and 1.0 (Tokyo).

A comparison of the results (1) and (2) shows that there is a substantial complementarity effect for the short vowel from onset and coda in Chonnam, to the extent that the overall syllabic durations of the short and long series come out equal. This shows a typical characteristic of syllable-timed languages. Tokyo, on the other hand, maintains the distinction between the short and long series both in (1) and (2), a characteristic of mora-timed languages. These results suggest that the long vowel in Chonnam is not a bimoraic long vowel like the one in Japanese but is an expression of prominence.

Important to the current discussion, the results in (1) suggest that the vowel lengthening in Chonnam is not a universal phonetic phenomenon automatically occurring at the phrase initial position (cf. Keating, Cho, Fourgeron, and Hsu 1998): it is sensitive to the metrical prominence as well as the phrasal position, hence there is no lengthening effect for the unstressed initial syllables. Therefore, the phenomenon belongs to the phonetic module of the grammar before being subjected to universal phonetics.

Another piece of phenomenon that supports the argument of this study can be found from the laryngeal effect of Korean, which is known to be restricted to the initial syllable of a word (Kim 1965). Ko (1999) has shown that this F0 boosting effect is a phonetic effect, contra Jun (1993). Her argument was based on the tonal realization of the vocative chant, where only the H associated with a metrically salient position reaches the highest point of the available F0 range as its target. If the laryngeal effect is phonetic as Ko argues, then it provides another case of the positional effect found in phonetic grammar.

In sum, I have provided evidence for a level where the output of phonology is subjected to language specific phonetic grammar, and have suggested that the positional effect which has been already recognized to be in action in phonology is also in effect at this level.

Effect of pitch in Japanese word recognition: Evidence from computational analyses of the Japanese lexicon

Kiyoko Yoneyama and Keith Johnson (Ohio State University, Columbus)

Recent work by Cutler and Otake (1999a) shows that Japanese listeners are sensitive to pitch information during word recognition in Japanese. This paper will report the results of two computational analyses using a Japanese lexical database (about 65,000 phonetically-transcribed words; Yoneyama, 2000) to see how pitch information might contribute to word recognition in Japanese.

The first analysis examined uniqueness points in Japanese (Marslen-Wilson, 1984; Marslen-Wison and Tyler, 1987). The concept of uniqueness points is the core of Cohort Theory. This theory predicts that many words in the lexicon may be unique before the end of the word. However, Luce (1986) found that the words which have uniqueness points are the longer words in the lexicon, and that only 39% of the words have a uniqueness point (frequency-weighted analysis). Luce concluded that Cohort Theory may be severely limited in scope due to structural properties of the mental lexicon.

If prosodic information such as pitch is used, the calculated uniqueness point would occur before the end of word in more words than if only segmental information is used (as in Luce, 1986). Table 1 shows the materials used in our four uniqueness-point analyses of Japanese based on the combinations of two variables (types of representation and processing units). Table 2 shows the cumulative probability of words that have a uniqueness point before the end of the word in four analyses. The results here show that the smaller processing unit yields a higher probability of uniqueness points. The results also show that the richer representation with pitch information yields a higher probability. Therefore, the results of the four analyses show that the highest proportions of words having uniqueness points before the end of the word were obtained if the processing unit was segment and the representation contained pitch information. This result predicts that if Japanese listeners can pay attention to smaller units and pitch information, their word recognition performance will be maximized. This result is consistent with Cutler and Otake (1999ab). Still, the probability that a word diverges from all the other words in the lexicon before word offset was only .48 in Japanese whereas it was .59 in Luce's study of English. The results of the present lexical analyses therefore predict that words are not usually recognized before their offset in Japanese, similar to the results that Luce (1986) showed for English.

The second analysis investigated lexical neighborhoods in Japanese. Luce and Pisoni (1998) calculated lexical neighborhood size using a computerized on-line English dictionary and conducted auditory word recognition experiments. Their findings suggest that neighborhood structure may play an important role in word recognition. We calculated the number of neighbors for each word to investigate the structure of the Japanese lexicon in terms of sound similarity.

The neighbors of a lexical item are usually defined by using the Greenberg & Jenkins (1964) phoneme substitution, deletion or insertion rule. Two analyses were conducted. Analysis 1 considered only segmental similarity (Greenberg & Jenkins, 1964; Luce and Pisoni, 1998). Analysis 2 considered both segmental and prosodic similarity. Prosodic similarity was based on a pilot study in which Japanese native speakers gave ratings for selected contrasting pitch contours. In general, the pilot results show that Japanese listeners treat two contours which have only one tonal difference by substitution, deletion or insertion as similar. Therefore, in Analysis 2, neighbors were calculated for each word in terms of two dimensions: tones and segments.

There are two main findings. First, prosodic information reduces the number of words which have at least one neighbor. If the word length is more than five, Analysis 1 yields more words which have at least one neighbor than Analysis 2 (Table 3). Second, prosodic information reduces the mean number of neighbors as a function of the word length as measured by segments, when the word was short. The effect was greatest if the words were very short (Table 4). These findings predict that if the frequent words tend to be short in Japanese, prosodic information does not contribute to reducing the number of words that have at least one neighbor whereas it does contribute to reducing the mean number of neighbors.

Together, our analyses on uniqueness points and lexical neighborhoods both show that prosodic information may in theory contribute to word recognition in Japanese.

References

- Cutler, A. & Otake, T. (1999a). Pitch accent in spoken-word recognition in Japanese. Journal of Acoustical Society of America, 105(3), 1877-1888.
- Cutler, A., & Otake, T. (1999b). Phonemic effects in spoken-word recognition in Japanese. Journal of Acoustical Society of America, 106(4), 2276.
- Greenberg, J.H & Jenkins, J.J. (1964). Studies in the psychological correlates of the sound system of American English. Word, 20, 157-177.
- Luce, P.A. (1986). A computational analysis of uniqueness points in auditory word recognition. Perception & Psychophysics, 39, 155-158.
- Luce, P., & Pisoni, D.B. (1998). Recognizing spoken words: The neighborhood activation model. EAR AND HEARING, 19: (1) 1-36.
- Marslen-Wilson. (1984). Function and process in spoken word recognition: A tutorial review. In H. Bauma & D. Bouwhius (Eds.) Attention and performance: Vol. 10. Control of language processes (pp. 125-148). Hillsdale, HJ: Erlbaum.
- Marslen-Wilson & Tyler, L.K. (1978). Processing interactions and lexical access during word recognition in continuous speech. Cognitive Psychology, 10, 29-63.
- Yoneyama, K. (2000). Structural Aspects of the Japanese Lexicon. Paper presented at the Speakers Series at Department of Linguistics, Ohio State University.

Data

Table 1: Materials for four analyses

	Processing units	
	Segments	Morae
Representation without pitch	a.a.m.o.N.d.o	a.a.mo.N.do
Representation with pitch	a1.a1.m.o1.N0.d.o0	a1.a1.mo1.N0.do0

(0=low pitch, 1=high pitch; pitch is represented at the end of each mora; Dots indicate segment/mora boundaries)

Table 2: Proportion of words That diverge before the end of the word across four analyses.

	Processing units	
	Segments	Morae
Representation without pitch	0.387	0.272
Representation with pitch	0.483	0.362

Table 3: Percentages of words which have at least one neighbor as a function of word length in segments.

Length	Without pitch	With pitch
1	100	100
2	100	100
3	100	99.9
4	99.9	99.3
5	99.5	97.6
6	96.7	92.1
7	81.0	73.1
8	55.9	46.2
9	33.8	25.8
10	29.2	22.7
11	13.3	9.1
12	12.0	8.6
13	6.4	3.1

Table 4: The mean number of neighborsas a function of word length in segments.

Length	Without pitch	With pitch
1	99.8	53.3
2	101.4	46.3
3	94.2	66.8
4	53.0	30.0
5	41.9	25.4
6	33.1	26.6
7	9.3	8.0
8	2.5	2.4
9	1.5	1.4
10	1.4	1.5
11	1.1	1.1
12	1.1	1.1
13	1	1

14	8.0	2.9	14	1.3	1.8
15	6.9	5.2	15	1	1
16	3.1	3.1	16	1	1
17	0	0	17	0	0
18	14.2	0	18	1	0
19	0	0	19	0	0
20	0	0	20	0	0
21	0	0	21	0	0
22	0	0	22	0	0

Phonetic control of the [oral] feature for the contextually nasalized French oral vowels

Véronique Delvaux (Free University of Brussels)

In this study, we present data about the amount of nasalization in contextually nasalized French oral vowels and we discuss results in connection with Kingston and Diehl's hypothesis about phonetic control (Kingston and Diehl, 1994).

The selected dependent variable, i.e. the mean proportion of nasal to total airflow and volume, varies significantly according to the oral vowel considered. More specifically, proportional nasal airflow (PNA) of oral vowels that have a phonological nasal counterpart is nearly twice as small as PNA of other oral vowels. We examine some constraints that could account for the observed variation. The lack of a satisfactory explanation leads us to consider whether this phonetic variation in the implementation of the feature [oral] could be due to an active control of velum activity.

Eight native French speakers (from Belgium) took part in the experiment. Their task was to read lists of words containing the stimuli. Nasal and oral airflow were recorded with the Physiologia Workstation. Corpus contained 198 items with a NVN, NV or VN structure where N is one of the three French nasal consonants and V one of the ten French oral vowels. The PNA variable has the advantage of neutralizing the differences due to gender as well as the effects of changes in overall intensity.

The three main results are the following : PNA varies significantly according to (1) nasal consonant context type, (2) existence (or not) of a phonemic nasal counterpart to the oral vowel, and (3) vowel tongue height.

As regards context, progressive nasalization was found on average to be five times as large as regressive nasalization. Oral vowels preceded and followed by a nasal consonant are even more nasalized. However, an identical pattern occurs within each context : oral vowels can be divided into two groups : [a,E,9,O], the vowels that have a phonemic nasal counterpart [a~,e~,9~,o~], have a much smaller and steadier PNA than the remaining [e,2,o,i,y,u] (see Table 1).

Because of the well-known diachronic evolution of French nasality, high nasal vowels do not exist in this language. Thus, the "nasal counterpart" variable interacts with the tongue height variable. Nevertheless, the relationship between PNA and tongue height is not linear. In the first group, the lowest vowel /a/ has consistently more PNA than the three low-mid ones whereas within the second group (vowels without phonemic nasal counterpart), high-mid vowels are less nasalized than high ones (see Table 2).

Actually, high-mid vowels show PNA values much closer to low-mid than high vowels. This should not be surprising since the following facts are taken into account: (1) in most languages, mid vowels (either high-mid or low-mid) have a common and typical behavior with respect to nasalization, (2) in standard French, the phonological contrast between high-mid and low-mid vowels is neutralized in final position; most of the time, this is not true for Belgian speakers but there happens to be more variability in the realization of these minimal pairs in this specific context. In our data, the context variable interacts significantly with the relationship between high-mid and low-mid vowels. Finally (3), at least for the O/o pair, the vocal tract configuration of the nasal vowel (o~) is closer to the high-mid than the low-mid oral vowel in modern French. Both phonemes present rather similar values for PNA in our data while the E-e and 9-2 differences are much greater.

The results presented above do not provide an explanatory variable (i.e. a phonetic constraint) that would predict the observed variation in the amount of nasalization for contextually nasalized oral vowels with as much accuracy as does the phonemic nasal counterpart variable. Neither the coarticulatory source of nasalization type nor the features of the vocalic phoneme account for such a variation.

Thus, we propose to consider Kingston and Diehl's hypothesis about phonetic control. Control of the [oral] feature phonetic implementation in nasal context implies an active reduction of the amount of nasalization when a contrast with a nasal vowel has to be maintained, in order to maximize the acoustic and perceptual distinctiveness of [nasal] vs [oral] phonetic realization. Otherwise, control deliberately allows a greater amount of nasalization, which minimizes production effort.

Table 1 : PNA Means in %

	NV	VN	NVN
all	34,6	7	41,6
a	25,6	6	34,7
E,9,0	28,3	4	31,5
a,E,9,0	27,5	4,6	32,9
e,2,o	28,2	8,3	39,3
i,u,y	53,1	12,6	66,9
e,2,o,i,u,y	39,2	10,5	59

Table 2 : PNA Means in %

Segment	PNA Mean
9	16,5
E	17,7
O	20,2
a	22,4
9,E,O,a	19,9
2	22,8
o	22,1
e	25,9
2,o,e	23,6
y	38,4
u	42,3
i	52,3
y,u,i	45
2,o,e,y,u,i	35

References

- Kingston and Diehl, 1994. Phonetic Knowledge, Language, 70, 3, 1994, pp.419-453.

English vowel-liquid monosyllables: A case of trimoraic syllables

Abigail Cohn (Cornell University) and Lisa Lavoie (Harvard University)

In this study we bring together several sources of evidence--from phonology, speaker intuition, and phonetics--to investigate the structure of a subset of monosyllabic words in English, those with a vowel-liquid rime. We find that the behavior of these forms varies depending on the vowel at hand, as illustrated in (1) for /l/-rimes. Those with a lax vowel (a) or a low vowel (b) behave similarly to other vowel-consonant rimes, while those following a [tense, -low] vowel (c) or diphthong (d) have special characteristics.

(1)

a.	[lax, -low]	I, U, E, V	fill, full, fell, dull
b.	[+low]	{, a, O	pal, doll, fall
c.	[tense, -low]	i, u, e, o	peel, pool, pail, pole
d.	diphthong	aj, aw, oj	file, foul, foil

Based on a range of evidence, we conclude that these rimes are superheavy and argue that this is best captured by a trimoraic structure. We propose an Optimality Theoretic analysis and present phonetic measurements that both confirm our analysis and lend additional support to the fact that moraic structure is one of the organizing factors of phonetic duration. We thus conclude that English exhibits not only a light-heavy contrast (a structural contrast between one and two moras), seen in its stress system and word minimality effects, but also a contrast between heavy and superheavy, exhibited by the behavior of rime liquids when preceded by diphthongs and [tense, -low] vowels.

The grouping of vowel nuclei preceding liquids shown in (1) emerges from our own intuitions and is supported by the judgements of six subjects based on a syllable count questionnaire. While the cases in a&b were uniformly judged to be monosyllables, the patterns in c and even more strongly those in d were often judged to be heavier. The heavier monosyllabic judgements follow from the proposal that these are trimoraic. Some judged those in c&d to be disyllabic. This suggests that for some speakers and/or dialects these are reinterpreted as disyllabic, consistent with the cross-linguistic dispreference for trimoraic syllables.

Further evidence for the special status of the cases in c&d comes from the patterning of these forms in chanting. Canonical disyllabic forms require a consonantal onset for the second syllable, as in "Maya" [maj-j@]. In contrast the canonical monosyllables show hiatus: "Bill" [bI-Il], *[bI-jl]. The superheavy cases pattern like the disyllables: "Neil" [ni-jl], "Kyle" [kaj-jl]. Evidence from the comparative (2) suggests that the diphthong-liquid cases (d) pattern with the disyllables in requiring "more" rather than "-er" (usually reserved for monosyllables and disyllables ending in an [i] or syllabic consonant):

(2)

-er	faster, happier, simpler, tenderer	a. shriller, b. smaller, c. cooler
more	more complex, more blasé	d. more vile, more loyal

We account for this distribution of English monosyllables within Optimality Theory, assuming that the lax vowels are inherently monomoraic and the tense vowels and diphthongs bimoraic. A constraint requiring all content words to consist of at least two moras forces the first coda consonant after a lax vowel to bear a mora. A constraint requiring rime liquids to bear a mora accounts for the appearance of trimoraic syllables when an inherently bimoraic nucleus is followed by a liquid. In the [+low] cases, which include bimoraic vowels (such as [a]), a high ranking constraint prevents a following liquid from bearing a mora due to the high sonority of the low vowels.

Arguing that duration is derived from both mora and segment count, we predict that the trimoraic rimes will show longer overall duration than similar (bimoraic) rimes with non-liquids. In an acoustic study of /l/-rimes of two speakers of American English, we find systematic differences in duration, correlating with the proposed differences in mora count. Duration of /d/ is found to be independent of moraic structure, while Vl nuclei (as compared to V nuclei) have additional duration from the presence of /l/. Comparing, for example, low vowels (A) with diphthongs (AW), we find that the CV and CVd cases, both argued to be bimoraic, are closely parallel in duration [CV: A=139 ms, AW=143; CVd: A+d=126+52, AW+d=128+57]. However there is a systematic difference for the CVl and CVld cases, argued to be bimoraic for low vowels and trimoraic for diphthongs [CVl: Al=154, AWl=201; CVld:

Al+d=162+51, AWl+d=207+58]. These results strongly support the proposed analysis and offer insight into how moras and segments are integrated in the implementation of duration.

In summary, we explore the behavior of a class of monosyllables in English and find subtle but systematic differences in both phonological behavior and phonetic realization, supporting the conclusion that these syllables are monosyllabic, but trimoraic. The inherent instability of trimoraic syllables explains the variation in the realization of these patterns both in terms of speaker intuition and dialect variation. This work suggests that small but systematic differences in the behavior and realization of rime consonants follows from a moraic/non-moraic distinction.

Articulatory considerations in continuancy alternations

Lisa Lavoie (Harvard University)

Many languages exhibit alternations between voiced stops and their continuant counterparts (e.g. Basque, Catalan, Efik, Pennsylvania German, Panamint, Senoufo, Spanish, Tamil, Tatar, Tzeltal). Generally, the stops occur word- or utterance-initially and post-nasally, while the continuants occur intervocalically. The most prominent case of this type in the literature is Spanish. The status of Spanish voiced stop-continuant alternations has been the subject of numerous studies (e.g. Harris 1969, 1983; Lozano 1979; Florian 1985; Lipski 1987; Hualde 1991; Bakovic 1995). These studies have focused on the choice of underlying segment, precise characterization of the environment, and the rules or constraints responsible for the alternation. These studies all assume that the continuancy alternation is part of the phonology; its variability is accounted for with optional or variable rules. The earlier work is usually consistent with the actual historical development in assuming underlying stops with spirantization while later work may assume underlying continuants with hardening. These data pose a number of interesting questions. Why do the same alternations and environments occur in so many languages? How do these alternations fit into the broader scheme of consonant lenition or weakening? Is a stop or fricative more articulatory effort? Are the continuants fricatives or approximants? If the alternations are part of the phonology, what is the nature of the variability? Recent work on articulatory strengthening by Keating and her collaborators offers some insight as does work by Ohala on post-nasal hardening.

With the notable exception of Cole, Hualde & Iskarous's (1997, 1999) acoustic studies, most of this work has been based on impressionistic observation. Spanish obstruent alternations represent a case in which the phonetic data are essential for a complete picture. In addition to providing hard evidence of manner of articulation, the phonetic data protect the experimenter from the effects of categorical perception. The boundary between incompletely-closed stops and frictionless continuants or approximants is quite difficult to perceive. We have no phonemes that are incompletely closed stops; it is a very unstable category. While the acoustic data provide clearer information on these alternations, they cannot quantify degree of contact/closure, information which is available from articulatory studies.

I carried out parallel acoustic and articulatory (electropalatography) studies of Mexican Spanish. In the articulatory study, I studied coronal and velar obstruents in four different prosodic positions in real words and reiterant speech. As expected, the utterance-initial segments were realized as stops. These results closely parallel Fougeron & Keating's (1997) and Fougeron's (1998) findings of articulatory reinforcement at the edges of prosodic domains in English and French. Articulatory reinforcement or strengthening was manifested as greater linguopalatal contact in segments in the initial positions of prosodic domains, with the greatest contact in utterance-initial position. I argue that the choice of underlying continuants with phonetic strengthening is the appropriate analysis for Spanish. The two phonetic effects which yield increased contact, domain-initial strengthening and post-nasal hardening, are responsible for the realization of the underlying continuants as stops.

Once the alternations are understood as resulting from underlying continuants, another consonant alternation that is not usually thought to be related assumes its place under the same analysis. Mexican Spanish palatal glides are realized as palatal affricates in word- or utterance-initial position. Even utterance-medially, the palatal glide is articulated with almost a complete seal. Just a bit more contact, attributable to articulatory strengthening, completes the seal, yielding a palatal stop, a segment that is often affricated (Keating 1988). Like the voiced stop-continuant alternations, the palatal glide-affricate alternation is attributable to the phonetics and should not be considered part of the phonology.

In contrast to these voiced segments, Mexican Spanish /s/ does not participate in this variation. Sibilants require very precise articulation (e.g. Ladefoged & Maddieson 1996, Dagenais, Lorendo & McCutcheon 1994). Fougeron (1998) found that /s/ shows the least amount of variation in contact in French. Lavoie (2000) found almost no variation in contact pattern for /s/ in English or Spanish. The Spanish voiced obstruents and /s/ illustrate the range of possible phonetic variation. The voiced obstruents allow more phonetic variation with the result that they actually move into a different perceptual category, that of non-continuant. The different behavior of these segments can be captured in Keating's (1988) window model by positing a very small window for /s/ but a larger one for the voiced obstruents. The larger window along the phonetic dimension of contact opens up the possibility that variation will yield the percept of distinct allophones. Besides relating to the question of whether stops or fricatives require more precision and/or effort, these data all ultimately bear on the question of how phonetic variation that yields categorical results fits into a view of the phonetics-phonology interface which considers gradiency to be

phonetic but categoriality to be phonological (e.g. Cohn 1990, Keating 1996).

Constraints on prosodic phrasing in spontaneous speech - data from Swedish

Petra Hansson (Lund University)

This paper makes some preliminary hypotheses about the hierarchy of constraints on prosodic phrasing in spontaneous speech. Within Optimality Theory, phenomena such as prosodic phrasing are explained through the interaction of constraints which determine what prosodic form an utterance is assigned given its syntactic and information structure. Among these are constraints that both align prosodic phrase structure with syntactic and information structure and constraints that delimit the prosodic phrases' size and tonal content.

For English, Selkirk (In press) has proposed a hierarchy in which the constraints WrapXP and AlignXP occupy the same rank, meaning that an utterance may either be contained within one prosodic phrase or have as many prosodic phrases as there are maximal categories. Two size constraints are also proposed as well as a constraint AlignFocus which calls for the right edge of a focus constituent in information structure to be aligned with the right edge of a prosodic phrase.

The constraints on output representations are hypothesized to be universal. Languages differ only in the ranking order of the constraints. Therefore it is interesting to investigate in what way the above constraints can be assumed to interact in Swedish. Whereas previous studies on constraints on prosodic phrasing have been limited to read stimuli, we have chosen to use spontaneous dialogues as data.

Spontaneous speech is typically less structured syntactically than written discourse. It contains fragments and there is little subordination. A first observation about the prosodic phrasing in our data is that very few utterances have internal prosodic phrase boundaries, suggesting that WrapXP be ranked higher than AlignXP. An exception are utterances with several focal accents and utterances containing speech repairs. Heeman (1997) has shown for English that the end of a disfluency reparandum (the interruption point) is often accompanied by a disruption in the intonation contour. Intonational disruptions and boundary tones are also found in speech repairs in Swedish. When formulating constraints on prosodic phrasing in spontaneous speech, one needs to include speech repair phenomena. It is thus proposed that the AlignXP constraint also calls for the right edges of incomplete XPs to be aligned with the right edges of prosodic phrases (PPh) (see e.g. 'till' ('for') in (1)).

According to the Lexical Category Condition (Truckenbrodt, 1999), constraints relating syntactic and prosodic structure apply to lexical elements and their projections, but not to functional elements and their projections. However, in spontaneous Swedish, demonstrative pronouns used to point out a referent in the external world are often followed by a phrase boundary (see e.g. 'ett sådant' ('one of these') in (1)).

(1)
(jo du "måste ha ett 'sådant "kort till)PPh (ett 'sådant)PPh (till "bägge de 'här va)PPh
'(yes you need to have one of these cards for) (one of these) (for both of these [tickets])'

In order for the phrasing in (1) to be considered optimal, we also need to formulate constraints which delimit the size and tonal content of a prosodic phrase. First, a constraint BinMax(PPh) is formulated to prevent utterances with more than two focal accents (indicated with ") to be 'wrapped' into one single prosodic phrase. It would allow the AlignXP constraint to align 'kort', 'till', 'sådant' and 'va' with prosodic phrase boundaries. Second, a constraint Min(PPh) is introduced to determine the minimal accentual content of a prosodic phrase. Since a prosodic phrase minimally contains one prosodic word and the defining feature of the prosodic word is that it contains an accent (indicated with '), Min(PPh) calls for a PPh to consist of at least one accent. Min(PPh) is the highest-ranked constraint in our hierarchy. It prevents 'till' from constituting a prosodic phrase on its own. The following hierarchy is proposed for spontaneous Swedish: Min(PPh)>>BinMax(PPh)>>WrapXP>>AlignXP.

In this paper an attempt is made to determine whether empirical evidence can be found for theoretical claims made about a number of universal constraints on prosodic phrasing in spontaneous speech. Evidence for the presence and interaction of syntactic and size constraints in Swedish has been found. However, no evidence to support the AlignFocus constraint has been found so far. In order for the constraint hierarchy to apply to spontaneous speech, we have proposed that AlignXP allows incomplete XPs and demonstrative pronouns to be aligned with prosodic phrase boundaries.

References

- Heeman, Peter. 1997. Speech Repairs, Intonational Boundaries and Discourse Markers: Modeling Speakers' Utterances in Spoken Dialogue. Doctoral dissertation.
- Lambrecht, Knud. 1994. Information structure and sentence form. Topic, focus, and the mental representation of discourse referents. Cambridge Studies in Linguistics 71. Cambridge: Cambridge University Press.
- Selkirk, Elisabeth. In press. The Interaction of Constraints on Prosodic Phrasing. In Merle Horne (ed.), Prosody. Theory and Experiment. Studies presented to Gösta Bruce. Dordrecht: Kluwer.
- Truckenbrodt, Hubert. 1999. On the Relation between Syntactic Phrases and Phonological Phrases. Linguistic Inquiry 30, 219-255.

Phonological categories and their phonetic correlates: The case of retroflexes

Silke Hamann (University of Utrecht)

This paper considers the acoustic and articulatory correlates of the phonological class of retroflex consonants. It presents the preliminary results of a series of experiments investigating the acoustic characteristics of retroflexes and their stability for categorizing segments. The fact that retroflexes constitute a phonological category is undoubted and attested by several phonological rules referring to them. One of these rules is for example the Nati rule in Sanskrit (Schein & Steriade 1986) which spreads the features of the retroflexed rhotic and voiceless fricative to a following /n/. Despite this fact, there is no agreement in the phonological description of retroflexes by means of features. Proposals differ widely, ranging from the traditional, SPE-oriented [-distributed, -anterior] to the monovalent [COR, apical] used in Feature Geometry (e.g. Gnanadesikan 1994 or Hamilton 1993). All features used in these descriptions are defined exclusively in terms of articulation, which appears to be the wrong approach. As several recent phonetic studies show, there is no consistency in the articulation of retroflexes. Dixit (1990) e.g. proves for retroflex stops in Hindi, that the place of articulation can vary from alveolar ridge to the midpalatal region of the tectum and the articulator from the underside of the tongue tip to the blade. Furthermore, the degree of retroflexion can differ across subjects and is to a large extend dependent on the context of the sound (see also Krull et al 1995). Such an amount of variation in the articulation of retroflexes makes a precise correlation of articulatory definition and phonological features impossible. Both the place of articulation and the articulator overlap largely with that of other coronal categories. For this reason, Jakobson, Fant and Halle's (1952) suggestion of classifying sounds by acoustically defined features was reconsidered and applied to retroflexes. In the experiments undertaken, three Norwegian and three Tamil speakers were recorded reading 10 minimal pairs including retroflex and non-retroflex coronals, embedded in a trigger sentence. The analyses included comparisons of the formant transitions over time, the length of closure time (for plosives) and actual formant frequencies.

The results show that the zero in the spectrum of retroflexes which is due to their sublingual cavity causes a lowering of the third formant. An additional formant which is also introduced by the sublingual cavity results in a greater bandwidth of the second formant and sometimes produce a movement towards the third formant. These two criteria, i.e. the lowering of F3 and the clustering of F2 and F3, where already included in Jakobson, Fant and Halle's definition of the feature flat, which referred not only to retroflex sounds but also to lip-rounding and velarization.

Contrary to Stevens and Blumstein (1981)'s assumption, the data found demonstrates that the formant frequencies of a specific time are not sufficient to distinguish retroflexes from other coronals. The transitions of the second and third formants into the retroflex consonant only yield a stable criterion, a result that is in accordance with Krull et al's findings. In comparison to Jakobson et al's feature system, the author proposes to retain the articulatory definitions of the remaining coronal consonants, as they are phonetically correct and sufficient. The resulting phonological feature system therefore contains both articulatory and acoustically defined features, e.g.:

	dental	alv	retroflex	palato-alveolar
CORONAL	+	+	+	+
[apical]	-	+	+	-
[flat]			+	
[posterior]				+

References

- Chomsky, Noam and Morris Halle (1968) The Sound Pattern of English. New York: Harper and Row.
- Dixit, R. Prakash (1990) "Linguotectal contact patterns in the dental and retroflex stops of Hindi." Journal of Phonetics 18: 189-201.
- Gnanadesikan, Amalia (1993) "The Feature Geometry of Coronal Subplaces." University of Massachusetts Occasional Papers in Linguistics 16: 27-67.
- Hamilton, Philip James (1996) "Phonetic constraints and markedness in the phonotactics of Australian Aboriginal languages." Toronto working papers in linguistics.
- Jakobson, Roman, Gunnar Fant and Morris Halle (1952) Preliminaries to Speech Analysis: The Distinctive Features and their Correlates. Cambridge: MIT Press.

- Krull, D, B. Lindblom, B.-E. Shia and D. Fruchter (1995) "Cross-linguistic aspects of coarticulation: an acoustic and electropalatographic study of dental and retroflex consonants." Stockholm: Proceedings of the ICPhS XIII; 436-439.
- Stevens, Kenneth and Sheila Blumstein (1981) "The search for invariant acoustic correlates of phonetic features." In: P.D. Eimas & J.L. Miller (eds.) Perspectives on the study of speech. Erlbaum: New Jersey; 1-35.

Intelligibility of time-compressed speech and the role of word prosody

Esther Janse (University of Utrecht)

When speech is time-compressed, e.g. for the purpose of fast playback of voice mail messages, compression is normally performed in a linear way. Duration analyses have shown, however, that when people speak faster, durations of unstressed syllables are reduced more, relatively speaking, than durations of stressed syllables (Peterson & Lehiste 1960, Port 1980, Van Santen 1994). Obviously, when talking fast, speakers do not compress speech segments in a linear way. This raises the question whether the intelligibility of time-compressed speech would improve if its temporal organisation would be closer to natural fast speech. In spoken-word processing listeners match the incoming signal with word templates stored in the mental lexicon. Stress patterns are part of these templates. When segmental information deteriorates due to an extreme speech rate, prosodic patterns supposedly become relatively more important to intelligibility. Now we may hypothesise that listeners profit from a linear transformation because they can map the incoming signal onto the stored word template with a rather simple transformation. Alternatively, mapping the input onto a template might be easier if the duration relations within the word are more similar to those found in natural fast speech, where the temporal consequences of the stress patterns are made more pronounced than in normal speech rate.

The relative importance of stress-pattern related word prosody for intelligibility of time-compressed speech was studied in two perception experiments. In the first experiment intelligibility was established of words embedded in spoken sentences which were either time-compressed in a linear way or time-compressed in a way similar to what speakers do when talking fast. It appeared that the words in the linear compression condition were better intelligible than the words which were compressed in the way speakers naturally do.

In the second perception experiment, we tested the hypothesis that low intelligibility in time-compressed speech is mainly due to unstressed vowels becoming too short to be perceptible. We compared the intelligibility of linearly compressed speech with the intelligibility of speech in which the duration difference as a result of lexical stress was removed before the speech was compressed. We expected that the intelligibility would be better in the linear compression condition because in the other condition there was a striking mismatch between word prosody in the stimulus and word prosody in the mental word templates. We also tested the intelligibility of non-words in the two compression conditions because in non-words there obviously cannot be a mismatch between stimulus and word templates. As expected, at similar compression rates, intelligibility of real words was worse in the condition where stress-related duration differences were removed before compression. Also, as predicted, manipulation of vowel durations was less harmful to the intelligibility of non-words.

The results of these two perception experiments show that word prosody contributes to recognising words in highly time-compressed speech. However, applying selective compression in the same way as speakers do when talking fast does not appear to be beneficial for the intelligibility of compressed speech. We suggest this may be related to the fact that heavy coarticulation in naturally produced fast speech cannot be properly imitated by compression. In such natural fast speech, the duration of an unstressed vowel can often not be detected at all, yet heavy coarticulation gives some auditory impression of the otherwise missing vowel. This means that the results of our first perception experiment do not tell us whether the natural timing of fast speech is or is not beneficial for the intelligibility of fast speech. Imitation of natural timing might still be beneficial in case it is accompanied by natural coarticulation.

References

- Peterson, G.E. & Lehiste, I. (1960), Duration of syllable nuclei in English, Journal of the Acoustical Society of America 32 (6), 693-703.
- Port, R.F. (1981), Linguistic timing factors in combination, Journal of the Acoustical Society of America 89 (1), 262-274.
- Van Santen, J.P.H. (1994), Assignment of segmental duration in text-to-speech synthesis, Computer Speech and Language 8, 95-128.

The perception of geminates in Hindi

Manjari Ohala (San Jose State University) and John J. Ohala (University of California, Berkeley)

This study attempted to build on some of the pioneering experimental work of Hankamer and Lahiri (1989) (H&L) on the phonetic and phonological character of geminates stops. The duration of the stop closure was the primary cue differentiating geminate vs. singleton: lengthening singletons' closure and shortening geminates' closure duration succeeded in transforming one category into the other. However, these two curves specifying the perceptual differentiation between singletons and geminates as a function of closure duration did not overlap completely: there was a range of durations where the geminates were still predominantly heard as geminates even when they had durations which still cued non-geminate in the lengthened singletons. Similar results were first reported by Lisker (1957) in a study of the cue value of stop closure duration for the intervocalic [+/-voice] distinction in English stops. We refer to this as the "duration gap". H&L concluded that some secondary cues in addition to the primary cue of closure duration were responsible for the duration gap. However, their examination of a number of acoustic cues (duration of the preceding vowel, VOT, duration of the vowel offramp of the first syllable, RMS value of the burst as well as of the entire second syllable, etc.) yielded inconclusive results. They did not find any of these possible secondary cues systematically related to the duration gap. They conceded that there are perhaps a combination of cues "...each by itself too subtle for our measurements to detect." We sought to extend their study in a search for possible secondary cues to the geminate-singleton distinction in stops.

The language used for our study was Hindi, which, similar to Bengali, has a intervocalic geminate/singleton contrast. Along with seeing if Hindi also yielded results which show the duration gap, we sought to localize whether these secondary cues might reside in the preceding or the following vowel. Our speech material was minimal pairs of geminate vs. singleton voiceless unaspirated stops at the bilabial and velar places of articulation uttered in a frame sentence by three male native speakers of Hindi: /paka/ 'ripe', /pak:a/ 'firm'; /Tapa/ 'make someone jump over', /Tap:a/ 'a type of song' ('T' = voiceless unaspirated retroflex stop). The words were extracted from the frame and digitized at 16kHz. We then created the following types of stimuli: original geminate or singleton intact; original word's stop closure duration shortened or lengthened, respectively, in four 20 msec increments; the two halves of the original geminate or singleton cross-spiced with these same variable closure durations. Thus tokens had the shape (each with variable closure durations): /paka/, pak:a/, /pak + k:a/, /pak: + ka/. The object was to try to localize whether the secondary cues were predominantly in the preceding or following vowel. The resulting 120 tokens were randomized and played to native Hindi speaking listeners over ear phones individually in a relatively quiet environment. Preliminary results based on 5 subjects show results similar to Bengali:

- Duration is the primary cue for the geminate vs. singleton distinction in Hindi stops.
- However, other secondary cues are also implicated since the 'duration gap' was occasionally found on some minimal pairs.
- These secondary cues seem to be based sometimes on the preceding vowel and sometimes on the following vowel, depending on the given minimal pair and the speaker.

If the results from additional subjects corroborate these results it would indicate a multiplicity of secondary cues spread more globally through the word. In as much as there are secondary cues influencing listener's judgments, it requires some modification of the claim within current phonological theory that for the perception of geminates (unlike other consonant types) the timing tier is the only relevant factor (Lahiri and Marslen-Wilson 1992).

References

- Hankamer, Jorge, Lahiri, Aditi, & Koreman, Jacques. 1989. Perception of consonant length: voiceless stops in Turkish and Bengali. Journal of Phonetics, 17, 283-298.
- Lahiri, Aditi & Marslen-Wilson, William. 1992. Lexical processing and phonological representation. In: Gerard J. Docherty and D. Robert Ladd (eds) Papers in Laboratory Phonology II, Gesture, Segment, Prosody. Cambridge: Cambridge University Press.
- Lisker, L. 1957. Closure duration and the intervocalic voiced-voiceless distinction in English. Language 33 , 1, 42-49.

Quantity contrasts: Production, perception and frequency in Finnish and Japanese

Katsura Aoyama (University of Hawai'i at Manoa)

Purpose of the study

This study investigates how quantity contrasts are processed in two different languages, Finnish and Japanese. It was found that quantity contrasts occur more frequently in Finnish than in Japanese, and the difference in frequency seems to be reflected in both the production and perception of the contrast.

Experiment 1.
Single/geminate nasal boundary in Finnish and Japanese: Production

Method

A total of ten Finnish speakers and ten Japanese speakers participated. The target words were two names that differ only with respect to the length of the medial nasal, Hana and Hanna. A total of 240 tokens were collected and analyzed. Wide-band spectrograms were produced for each word, and the durations of each segment except the initial [h] were measured.

Results

The distinction between single and geminate nasals appears to be clearer in Finnish than in Japanese. The productions of geminate nasals were similar between the two languages both in terms of the mean absolute durations (178 ms. in Finnish and 172 ms. in Japanese; this difference is not significant) and in their proportions in the word (49.3% in Finnish and 50.8% in Japanese on average). However, the durations of single nasals were significantly shorter in Finnish than in Japanese (62 ms. in Finnish and 68 ms. in Japanese) ($t = 3.015$, $p < 0.005$). The nasal portion in [ana] was 23.8% in Finnish, while it was 32.8% in Japanese.

		Vowel 1		Nasal		Vowel 2		Whole word
		(ms)	%	(ms)	%	(ms)	%	(ms)
F:	[ana]	90	34.5%	62	23.8%	109	41.8%	261
J:	[ana]	53	25.5%	68	32.8%	87	41.8%	208
F:	[anna]	102	28.3%	178	49.3%	81	22.4%	361
J:	[anna]	77	22.8%	172	50.8%	90	26.6%	338

Table 1. The comparison of the vowels and the medial nasal in [ana] and [anna].

Experiment 2.
Single/geminate nasal boundary in Finnish and Japanese: Perception

Method

A total of 20 Finnish speakers and 25 Japanese speakers participated. A hana-hanna continuum with ten incrementally different stimuli was prepared in each language by deleting twelve-millisecond increments from the original hanna utterance. The prepared audio-stimuli were randomly presented to the subjects. Subjects categorized the stimuli as either hana or hanna by pressing the appropriate buttons. The set of stimuli was presented ten times; a total of 100 stimuli were presented to each subject.

Results

The incremental changes in duration were perceived categorically by both Finnish and Japanese speakers. The categorical boundary between single and geminate nasals seems to be similar, both in terms of the duration of the medial nasal (105 ms. in Finnish and 107 ms. in Japanese), and the proportion of the nasal in the word (36.7% vs. 36.9%). However, Finnish speakers have a narrower 'bandwidth' for their categorical boundary than Japanese

speakers. Six stimuli out of ten were categorized as either hana or hanna more than 95% of the time in Finnish while only five stimuli out of ten were categorized as such in Japanese. Thus the bandwidth of the categorical boundary was approximately 48 ms. in Finnish, while it was approximately 60 ms. in Japanese.

Frequencies of long vowels and geminate consonants in Finnish and Japanese

Method

The frequency of occurrence of quantity contrasts was counted in written texts in Finnish and Japanese. Data were prepared in CHAT format and phoneme frequency was calculated using the tools provided by the Child Language Data Exchange System (MacWhinney 1995).

Results

It was found that both long vowels and geminate consonants occurred more frequently in Finnish than in Japanese: the ratios between short and long vowels, and single and geminate consonants were 9.9:1, and 9.2:1 respectively in Finnish, while they were 15.0:1, and 19.9:1 in Japanese. This means that, at least in these data, quantity contrasts were almost twice as frequent in Finnish as in Japanese.

Conclusion

The frequency count revealed that geminates and long vowels occur more frequently in Finnish than in Japanese, which suggests that quantity contrasts are potentially more salient in Finnish than in Japanese. It seems that the differences in frequency of the quantity contrast between these two languages are reflected in both native speakers' production and perception; in production, the absolute duration of the single nasal [n] was shorter in Finnish than in Japanese, and the proportion of the nasal was smaller in the word (h)ana in Finnish than in Japanese. In perception, although categorical perception was observed in both languages, Finnish speakers had a narrower bandwidth of the categorical boundary than Japanese speakers, and slope of the categorization was sharper in Finnish than in Japanese. The high frequency of quantity contrast may lead speakers to sharpen the boundary between contrasts in quantity in phonological processing.

References

- MacWhinney, Brian. 1995. The CHILDES Project. Hillsdale, NJ: Lawrence Erlbaum.

Phonological variation as evidence for lexical representation of homonyms

Daniel Jurafsky, Alan Bell, and Cynthia Girand (University of Colorado, Boulder)

The role of the lexicon has been the focus of a wide variety of research in phonology. In this paper we propose to use phonological variation as a tool to shed light on the nature and constituency of the lexicon. A key problem in building a complete model of the lexicon is understanding the complex relationship between semantically-defined lexical entries (`lemmas' in the terminology of Levelt 1983), underlying phonological forms (`lexemes'), and surface forms (`wordforms'). Many wordforms, for example are homonymous; the same form is associated with multiple senses or lexical categories. The wordform `to', for example, is ambiguous between an infinitive marker ("we had `to' do it") and a preposition ("I would have gone `to' the store"). Our earlier work (Jurafsky et al. 1998) argued that the relationship between the (semantic) lemma and (underlying phonological) lexeme level was also complex. Words like `the', `a', and `to' seemed to have multiple allomorphs (for example the lemma `the' seemed to have two allomorphs THEE and THUH), the selection of which is dependent on many complex speaker variables.

To what extent are the multiple senses or categories of each of these words distinct lemmas? Earlier studies (Levelt 1983, Roland and Jurafsky 1999) have argued that the lemma (a semantically or syntactically-defined unit) plays a more important role in production than the lexeme or wordform (phonologically or orthographically defined units). If this is true, and if the preposition `to' and the infinitive marker `to' are indeed different lemmas, we could expect to find some effects in lexical production which distinguish between them.

Our proposal is to look at differences in the surface phonology and phonetics of these words in context to provide evidence for distinct lemma structures. Linguists have long noticed that phonological features such as lexical stress play a role in distinguishing for example nouns (OBject) from verbs (obJECT). We propose that phonetic or phonological reduction or lenition is a variable that can play a similar role in differentiating different lemmas. Ladefoged (1993), for example, noted anecdotally that the complementizer `that' is more likely to be reduced than the pronoun `that'. Vielleux and Shattuck-Hufnagel (1998) noted that the preposition `to' is pronounced differently than the infinite marker `to'. These results, however, do not control for the many variables that are known to affect pronunciation variation. Our work extends these very preliminary studies to look at the general role that syntactic categories play in lexical representation and production.

Our study is based on the Switchboard corpus of 2430 telephone conversations between strangers, collected in the early 1990s (Godfrey et al. 1992). Approximately four hours (38,000 word tokens) of this speech was phonetically hand-transcribed by students at UC Berkeley (Greenberg et al. 1996).

We hand-coded four frequent English function words from this database for their syntactic categories: `that', `to', `of', and `you'. For example, we investigated four different `thats': (`complementizer', `pronoun', `determiner', and `relative pronoun') and two different `tos' (`preposition', `infinitive marker').

The phonological variable we used to study these parts of speech was the reduction or lenition of the word's pronunciation in conversational speech. Our previous work (Jurafsky et al. 1998, Bell et al. 1999) studied a number of ways to measure reduction, as well as a wide variety of factors which influenced reduction. Based on this work, we coded three measures of reduction on the words `that', `to', `of', and `you':

- length in milliseconds
- reduced vowel
- deletion of final consonant (for `that' and `of')

We and others have shown that many factors influence reduction in function words, including: the speaker's rate of speech (in syllables per second), whether the speaker was having planning problems (as indicated by neighboring disfluencies), the position of the function word in the utterance, the segmental context, the predictability of the function word, and sociolinguistic factors such as age and sex. We controlled all our data for these factors and then tested if there was an effect of lexical category on reduction. In other words, for each of our three dependent measures of reduction (vowel quality, total word length, and final obstruent deletion), we checked whether different lexical categories of `that', `to', `you', and `of' behaved differently.

We did indeed find differences between the surface pronunciations of these forms, even after controlling for these factors and others. For example we found that the complementizer and relative pronoun categories for `that' were significantly more likely to have final /t/ deleted than were determiners. For the word `to', the infinite marker was

much more likely to have a reduced vowel than the preposition. These and other differences across wordforms in our results suggest that these words do consist of distinct lemmas, and confirm the key role of the lemma in lexical production.

Phonological encoding: In search of the lost syllable

Niels Schiller (Max Planck Institute for Psycholinguistics), Albert Costa (Harvard University), and Angels Colomé (University of Barcelona)

Phonological encoding (PE) in speech production refers to the retrieval of word forms from the mental lexicon. One question to ask is "What are the relevant phonological units speakers retrieve during PE?". Syllables may be a good candidate for playing an important role in PE. Indeed, the two most influential models of PE assume the existence of syllabic units (Dell, 1986, 1988; Levelt, 1989; Levelt, Roelofs, & Meyer, 1999). Despite this general agreement, these models widely differ in the status of syllabic units in PE. In Dell's (1986, 1988) model, for instance, the phonological composition of a word is already syllabified in the lexicon. In contrast, Levelt's model assumes that word forms are not syllabified in the lexicon; instead syllables are compu ted on-line during a syllabification process.

In spite of the experimental efforts devoted to the study of the role of the syllable in PE (e.g., Ferrand, Segui, & Grainger, 1996; Ferrand, Segui, & Humphreys, 1997; Levelt, & Wheeldon, 1994; Schiller, 1998, in press; Schiller, Meyer, & Levelt, 1997; Sevald, Dell, & Cole, 1995; Treiman, & Danis, 1988), the evidence in favor of such a unit is still scarce and sometimes contradictory. The most compelling experimental on-line evidence for the existence of syllables in PE is the so-called syllable priming effect obtained by Ferrand et al. (1996, 1997) for French and English. For instance, using a masked priming paradigm, Ferrand et al. (1997) found that English words starting with a CV syllable, such as "tomato", were named faster when preceded by a CV prime (i.e. "to") than when preceded by a CVC prime (i.e. "tom"). These findings, however, have not been replicated by Schiller (in press). In the present study we bring more experimental evidence regarding the syllable's role during PE in English.

Two experiments are reported that tested whether syllables can be primed speech production using a (masked) priming paradigm. In Experiment 1, we presented masked syllable primes for 45 ms. In Experiment 2, primes were presented for either 45 ms or 105 ms under unmasked conditions. The visibility of the prime was increased to ensure full processing of the prime--a prerequisite for observing syllabic effects. In both experiments we tested three different SOAs, namely (200 ms, 0 ms, and +200 ms. The SOA manipulation was made to present the prime at the moment in time when syllables are supposed to be computed. Therefore, these two novel manipulations were supposed to increase the likelihood of getting a syllabic effect. In all three prime presentation conditions phonological priming effects were obtained. However, no evidence for a syllabic priming effect was found. Instead, at SOAs (200 ms and 0 ms, priming effects increased when the segmental overlap between prime and target was increased. This out come supports a segmental overlap account but contradicts the syllable priming hypothesis. We suggest that the evidence in favor of the syllable's role in PE (Ferrand et al., 1997) should be carefully reconsidered.

References

- Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. Psychological Review, 93, 283-321.
- Dell, G. S. (1988). The retrieval of phonological forms in production: Tests of predictions from a connectionist model. Journal of Memory and Language, 27, 124-142.
- Ferrand, L., Segui, J., & Grainger, J. (1996). Masked priming of word and picture naming: The role of syllabic units. Journal of Memory and Language, 35, 708-723.
- Ferrand, L., Segui, J., & Humphreys, G. W. (1997). The syllable's role in word naming. Memory & Cognition, 35, 458-470.
- Levelt, W. J. M. (1989). Speaking. From intention to articulation. Cambridge, MA: MIT Press.
- Levelt, W. J. M., Roelofs, A, & Meyer, A. S. (1999). A theory of lexical access in speech production. Behavioral and Brain Sciences, 22, 1-75.
- Levelt, W. J. M., & Wheeldon, L. (1994). Do speakers have access to a mental syllabary? Cognition, 50, 239-269.
- Schiller, N. O. (1998). The effect of visually masked syllable primes on the naming latencies of words and pictures. Journal of Memory and Language, 39, 484-507.
- Schiller, N. O. (in press). Single word production in English: The role of subsyllabic units during phonological encoding. Journal of Experimental Psychology: Learning, Memory, and Cognition, 26, xxx-xxx.
- Schiller, N. O., Meyer, A. S., & Levelt, W. J. M. (1997). The syllabic structure of spoken words: Evidence

from the syllabification of intervocalic consonants. Language and Speech, 40, 103-140.
- Sevald, C. A., Dell, G. S., & Cole, J. S. (1995). Syllable structure in speech production: Are syllables chunks or schemas? Journal of Memory and Language, 34, 807-820.
- Treiman, R., & Danis, C. (1988). Syllabification of intervocalic consonants. Journal of Memory and Language, 27, 87-104.

A phonetician's view of phonological encoding

Pat Keating (University of California, Los Angeles)

This talk reviews some of the ideas associated with "phonological encoding". First I consider some ways in which phonology can be considered a code. Then I review models of and evidence for phonological encoding as a distinct step in speech production. It turns out that any explicit model of phonological encoding raises every major issue faced by phonology and phonetics -- e.g. the nature of lexical representations, how contextual variants are computed -- so that the psycholinguistic evidence on these issues has implications beyond phonological encoding per se. The second part of the talk concerns the role of suprasegmental structure in phonological and phonetic encoding. I review work presented at previous LabPhon conferences on phonetic variation as a function of prosodic context. Initial position is a locus of phonetic strengthening but also of speech errors, presumably because of its high informational load.

Morpheme boundaries, word boundaries, and glottal stops

Janet Pierrehumbert (Northwestern University)

English, like many other languages, exhibits a variable process of glottal stop insertion on vowel-initial words such as "Anne". Previous work has shown that both the likelihood and the degree of glottalization depend on lexical and phrasal stress as well as on intonational phrasing. Thus, glottal stop insertion provides a classic example of a post-lexical allophonic process.

In this paper, I explore a lexical influence on glottal stop insertion by examining the interaction of morphosyntactic structure with speech style. Materials were created in which a V-V hiatus occurred between a function word and a content word (as in "The Interstate ...") or between a prefix and a stem (as in "reentry"). These materials were recorded in both a normal fluent reading style and a clear speech style.

Probabilities of glottal stop insertion were influenced by speech style, morphosyntactic status, and stress. Clear speech style, the presence of a full word boundary, and stress all favored insertion. Summary percentages are as follows:

Boundary Type	Stress-Clear	Stress-Normal	Unstress-Clear	Unstress-Normal
#	100%	80%	52%	8%
+	75%	20%	12%	0%

Data are analyzed on the assumption that the glottal stops in words such as "reentry" represent a paradigmatic allophonic influence of the stem ("entry") on its morphological relative. I compare the consequences of the observed patterns for two formal architectures: probabilistic OT and a structured lexical network model with a fast phonological preprocessor and a dual-route morphological analyzer. The best analysis includes some key insights from both frameworks.

Phoneme frequency in spoken word recognition

Danny Moates, Z.S. Bond, and Verna Stockmal (Ohio University, Athens)

Is the recognition of a spoken word a word-level process, or is it a product of the processing of the phonemes composing the word? The present study explores the latter by examining the effects of phoneme frequency on recognition.

van Ooijen (1996) introduced the word reconstruction task as a new method for exploring word recognition. Participants hear a nonword (e.g., /tibl/) which can be changed to a real word by substituting either a vowel or a consonant. For example, /tibl/ becomes table when changing the vowel from /i/ to /e/ and becomes feeble when changing the first consonant from /t/ to /f/.

van Ooijen found listeners changed a vowel more often than they changed a consonant when given a choice and the vowel changes were made more quickly. When listeners were limited to changing only a vowel or only a consonant, they were more accurate in making vowel changes and the most common error was to change a vowel when a consonant change was called for. Collectively van Ooijen labeled these effects "vowel mutability."

Cutler, Sebastian-Galles, Vilageliu, and van Ooijen (in press) have replicated the vowel mutability effect in both Spanish and Dutch, thus establishing the crosslinguistic robustness of the effect.

The present study explored the hypothesis that frequently used phonemes are recovered more accurately and more quickly in word recognition than are less frequently used phonemes. The study manipulated phoneme frequency in a word reconstruction task, predicting that high frequency phonemes would show fewer errors than low frequency phonemes.

Method

Participants

Fifty-four native speakers of English were randomly assigned to three randomizations of the stimulus materials.

Materials

Fifty-six nonwords were created. All could be changed into two different words by altering either one vowel or one consonant.

The phonemes which, when substituted correctly, converted a nonword into a word were the target phonemes. Fourteen vowel target phonemes were matched with 14 consonant target phonemes of similar phoneme token frequency. Token frequency is the frequency with which a phoneme occurs in conversational English. The token frequencies were selected from norms by Mines, Hanson, and Shoup (1978). Mean token frequencies for vowel and consonant target phonemes were equivalent, M vowels = 1.82, M consonants = 1.81.

The two sets of 14 target phonemes were divided into high token frequency pairs (M = 2.72) and low token frequency pairs (M = 0.90) which differed significantly in token frequency, $t(13) = 8.12$, $p < .000l$.

As controls, the vowel words and consonant words were balanced for word frequency, number of possible words, and the order of occurrence for vowel vs. consonant substitution. Also, the location of the phoneme to be changed was distributed approximately equally across the beginning, middle, and end positions of the nonwords.

Twenty-eight more nonwords were created for filler items plus 27 for practice items.

Three randomizations of the 84 nonwords (56 experimental items and 28 filler items) were constructed such that one third of the experimental nonwords and one third of the filler nonwords occurred in each third of the list. Each third of the list was used in one of the three substitution conditions: vowel substitution, consonant substitution, and free choice (either vowel or consonant substitution) so that each participant was exposed to all three conditions.

The nonwords were presented on audiotape at 10-sec intervals.

Procedure

The 54 participants were tested individually. Each heard one randomization of the 84 nonwords over headphones and gave a spoken response to each nonword into a tape recorder for later transcription. Each participated in the free-choice, vowel, and consonant conditions.

Within each randomization the order of conditions was counterbalanced. Three participants were thus assigned to each of the six possible orders of conditions for each of the three randomizations. Results

The most relevant comparisons were the contrasts between high and low frequency targets in the vowel and consonant conditions. Mean errors were significantly higher in the consonant condition than in the vowel condition, consistent with the vowel mutability effect. Errors for high frequency consonants, however, did not differ from errors for low frequency consonants. In the vowel condition, however, mean errors were fewer for high frequency vowels than for low frequency vowels.

Discussion

Because high frequency vowels show the fewest errors in word reconstruction they may be the most important contrbutor to the vowel mutability effect.

References

- Cutler, A., Sebastian-Galles, N., Vilageliu, O. S., & van Ooijen, B. (in press). Constraints of vowels and consonants on lexical selection: Cross-linguistic comparisons. Memory & Cognition.
- Mines, M. A., Hanson, B. F., & Shoup, J. E. (1978). Frequency of occurrence of phonemes in conversational English. Language and Speech, 21, 221-241.
- van Ooijen, B. (1996). Vowel mutability and lexical selection in English: Evidence from a word reconstruction task. Memory & Cognition , 24, 573-583.

Visual cues in the perception of speech timing and their implications for phonological structure

Haruo Kubozono (Kobe University)

This paper reports the results of a series of phonetic experiments carried out to explore the potential importance of visual information for the perception of Japanese prosody and its relevance to phonological structure. It is well known that visual information provides useful cues in the perception of speech (generally known as the `McGurk Effect'). However, research on bimodal perception has centered largely on segmental distinctions, and relatively little is known about the role of visual information in the perception of prosody. With this background, this paper is concerned with the question of whether or not native speakers of Japanese are able to accurately perceive temporal differences solely on the basis of visual information, specifically differences between monomoraic and bimoraic syllables: e.g. /sa.do/ `Sado Island' vs. /saa.do/ `third'; /ka.ko/ `past' vs. /kak.ko/ `bracket'; /ta.ni/ `valley' vs. /ta.nin/ `other people'; /kai.ga/ `picture' vs. /kai.gai/ `foreign land'. Since the mora is the basic temporal unit in the prosodic system of Japanese, these pairs of words involve differences in rhythmic structures as well as phonological length.

The experiment featured 30 test words consisting of 15 pairs of words such as the ones mentioned above. These 30 words were randomized and were read by a native speaker of Japanese. The readings were simultaneously recorded on digital video tape and cassette tape. Twelve hearing impaired adults and nine unimpaired adults--all native speakers of Japanese--participated. Each of the unimpaired speakers was presented with the following three kinds of stimuli: 1) visual information only (video playback without sound), 2) auditory information only (cassette tape playback), and 3) audio-visual information (video playback with sound). The hearing impaired subjects were given only the first kind of stimuli. After the playback of each test word, the test subjects were asked to appropriately choose between two similar words printed on an answer sheet for a total of 30 responses.

The experiment showed very interesting results. Overall, both groups of subjects yielded extremely good scores, even when presented only with visual stimuli. The unimpaired subjects took virtually perfect scores for auditory stimuli and AV stimuli alike, and attained very high accuracy (85%) in the visual test, too. The hearing impaired subjects were also able to distinguish between minimal pairs in the visual test, although no better than the unimpaired subjects. This indicates that visual information alone provides very useful cues for perceiving temporal differences such as those found in the test words.

On the other hand, it also turned out that visual information was not useful at all in distinguishing between pairs that contrasted in vowel length word-finally: e.g. /sa.do/ vs. /sa.doo/, /ru.bi/ vs. /ru. bii/. Both groups were able to correctly distinguish between these words only at a chance level, whereas they were perfectly good at distinguishing long/short vowels in word-medial positions: e.g. /sa.do/ vs. /saa.do/, /bi.ru/ vs. /bii.ru/. This suggests that vowel length is completely invisible in word-final positions although it is perfectly visible in word-medial positions. This type of position-dependent invisibility occurs only with words contrasting in vowel length (short vs. long), and not with words contrasting in terms of the presence or absence of a moraic nasal, diphthong, or geminate obstruent, e.g. /ta.ni/--/ta.nin/, /kai.ga/--/kai.gai/, /ka.ko/--/kak. ko/.

Position-dependent invisibility of vowel length is interesting by itself, but is even more interesting when considered in conjunction with a noted tendency whereby vowel length is neutralized in word-final positions in Japanese phonology (Kubozono, forthcoming). Thus, long vowels tend to be phonetically shortened word-finally in modern Japanese (e.g. /hon.too/ -> [honto] `true', /sen.see/ -> [sense] `teacher'); this shortening has become permanent in the history of some words (e.g. /tyoo.tyoo/ -> /tyoo.tyo/ `butterfly'). Note that this neutralization of vowel length is heavily dependent on position, i.e. it is restricted to word-final positions. The position-dependent neutralization of vowel length is observed in a more remarkable way in productive word formation processes such as compound truncation and the formation of zuzya-go (Japanese musicians' secret language involving metathesis). In these processes, word-final long vowels are shortened almost obligatorily, yielding three-mora outputs against the four-mora templatic form that would otherwise be produced.

In sum, the experiments described in this paper reveal the potential importance of visual information for the perception of moraic (temporal) structure in Japanese, on the one hand, and its relation to phonological patterning on the other.

References

- Kubozono, H. (forthcoming) "The syllable as a unit of prosodic organization in Japanese" to appear in F. Fery and R. van der Vijver (eds.) Structure and Typology of the Syllable, CUP.

A typological study of stress 'deafness'

Sharon Peperkamp and Emmanuel Dupoux (Laboratoire de Sciences Cognitives et Psycholinguistique, Paris)

During the processing of spoken language, utterances are coded phonologically in a prelexical representation. This representation is language-specific: it encodes only a small subset of those segmental and suprasegmental distinctions that are available at the universal phonetic level. Crucially, this limits our capacity to perceive distinct phonological properties of a foreign language, giving rise to phonological 'deafnesses'. For instance, French subjects, as opposed to Spanish subjects, have difficulties in distinguishing words that differ only in the location of stress (Dupoux et al. 1998). This is due to the fact that word stress predictably falls on the final syllable in French. Given this regularity, French speakers do not need to code stress in the prelexical representation. In Spanish, by contrast, stress is contrastive; hence, Spanish speakers need to code stress in this representation. Consequently, native speakers of French have problems distinguishing stress contrasts, while native speakers of Spanish do not.

In this paper, we test whether stress is encoded prelexically by native speakers of a variety of languages, taking the presence of stress 'deafness' as an indication of the absence of stress in the prelexical representation. Native speakers of French are widely known to have troubles with stress in foreign languages, whereas native speakers of Polish, another language with a fixed stress rule, do not seem to have the same problem. Therefore, it is likely that not all predictable stress systems give rise to stress 'deafness' in speakers dealing with such a system in their native language.

We argue that the patterns of stress perception by adult speakers shed light onto the way in which the prelexical representation is compiled during infancy. In particular, we assume that the prelexical representation is fixed during the first years of life, before full mastering of the native language. The infant's problem is thus to decide, on the basis of a limited amount of information, whether stress should be encoded or not. In the French case, this decision can be taken on the basis of a universal phonetic representation: given that all utterances end with a stressed syllable, infants can deduce that stress is not contrastive by paying attention to universal phonetic cues at utterance endings. In other languages with predictable stress, the regularity is harder to extract from the surface speech stream. Our aim is to determine which stress systems can be acquired by infants before the fixation of the prelexical representation. We distinguish four classes of languages, corresponding to four types of information that might or might not be available when infants decide whether stress should be coded prelexically. In languages of Class I, the stress rule can be acquired on the basis of a universal phonetic representation only; in languages of Class II, it can be acquired once other language-specific phonological information has been extracted; in languages of Class III and IV, it can be acquired only after function words and content words, respectively, can be segmented out of the speech stream.

Besides French, Class I includes Finnish; in this language, stress is word-initial and all utterances begin with a stressed syllable. Infants can extract the stress rule by focussing their attention on universal phonetic cues at utterance beginnings. An example of Class II is Fijian, which has a quantity-sensitive stress rule applying at the right edge of words. As soon as infants have acquired the syllable structure of their language and make a distinction between heavy and light syllables, they can infer the stress rule from utterance endings. We use Hungarian as an example of Class III. In this language, stress falls on the word-initial syllable, but - contrary to Finnish - utterances do not generally begin with a stressed syllable, due to the occurrence of unstressed function words. Infants therefore need to segment function words out of utterance beginnings in order to deduce the stress rule. Finally, Polish represents Class IV. Word stress in Polish falls on the penultimate syllable. Utterances therefore generally end with an unstressed syllable immediately preceded by a stressed one. However, this final trochaic pattern can be obscured not only by the presence of unstressed function words, but also by the presence of monosyllabic words. Thus, an utterance that ends with a monosyllabic word has stress on the final rather than on the penultimate syllable. Infants, then, can deduce the stress rule only once they segment speech into separate words.

We will present results of experiments assessing the perception of stress by adult native speakers of these four languages, and discuss the consequences for the acquisition of the prelexical representation. We conclude that cross-linguistic research of adult speech perception, when coupled with detailed linguistic analysis, can be brought to bear on important issues of language acquisition.

References

- Dupoux, E., C. Pallier, N. Sebastian, and J. Mehler 1997 A destressing 'deafness' in French? JML 36. 406-421.

Contextual variability, speaking style, and language background

Ann Bradlow (Northwestern University)

There is mounting evidence from cross-linguistic and laboratory-based production studies that contextual variability in the speech signal serves a communicative function. For instance, coarticulatory patterns are both language-specific(e.g. Manuel, 1990, JASA, 88, 1286-1298) and under talker control (Whalen, 1990, J. Phon., 18, 3-35), suggesting that contextual variability is planned and advantageous, rather than automatic and deleterious. Other studies showed that segments in prosodically strong positions resist coarticulation (Fowler, 1981, JSHR, 24, 127-139; Cho, 1999, ICPhS '99), suggesting that acoustic prominence and contextual variability are opposing forces. In order to gain further insight into the role of contextually induced variability for speech production and perception, we compared coarticulation across languages and across plain versus clear speech. Since the conversational-to-clear speech transformation involves listener-oriented articulatory adjustments, it is an effective window into underlying linguistic sound structure. Our primary questions were: (1) Is contextual variability minimized, maintained or exaggerated in clear speech? (2) Do the effects of hyperarticulation on consonant-vowel coarticulation vary across languages with vastly different vowel inventory sizes, e.g. English and Spanish? We hypothesized that coarticulation would be preserved even when talkers adjust their speech to maximize intelligibility. Furthermore, since vowel space expansion results in increased acoustic distinctiveness, it should be a feature of clear speech regardless of vowel inventory size. We were also interested in comparing the English and Spanish vowel spaces of monolingual and bilingual speakers to see if previously reported vowel space differences (e.g. Bradlow, 1995, JASA, 97, 1916-1924) are maintained or neutralized for bilinguals. We hypothesized that the coexistence of two vowel systems would result in subtle differences between the monolingual and bilingual vowel spaces.

English materials consisted of CV words where C = one of eleven consonants varying in place, manner and voicing, and V = /i/ or /u/. Spanish materials consisted of C1V1C2V2 words where C1 = one of ten consonants varying in place, manner and voicing, V1 = /i/ or /u/, and C2V2 = /so/. Target words were embedded in frame sentences, and produced in plain and clear speech (as if addressing hearing-impaired or non-native listeners.) Subjects were twelve American English monolinguals and nine Mexican Spanish American English bilinguals, yielding three data sets: English materials by English monolinguals (Eng-Eng), Spanish materials by Spanish-English bilinguals (Span-Span), and English materials by Spanish-English bilinguals (Span-Eng.) Vowel durations and formants (at 10 millisecond intervals) were measured. Current statistical analyses focus on /u/ in bilabial versus alveolar contexts.

Within data sets, results showed a significant main effect of speaking style on duration and F2: clear /u/ was longer and more peripheral in the vowel space than plain /u/. There was also a significant main effect of consonantal context: F2 was lower in bilabial than in alveolar contexts. Most importantly, the style-by-context interaction was not significant for any data set, indicating that the context effect was equivalent across speaking styles. This pattern held throughout the vowel duration. Thus, contextual variability was maintained in clear speech, suggesting that it is a systematic feature of both encoding and decoding mechanisms. Furthermore, contrary to views of stress as localized hyperarticulation (de Jong, 1995, JASA 97, 491-504) the maintenance of coarticulation in clear speech distinguishes it from prosodic strengthening, which involves coarticulatory resistance. Across data sets, there were no differences in the magnitude of the effect of style on /u/ location in the vowel space, indicating that vowel space expansion was a feature of clear speech regardless of vowel space "crowdedness." Additionally, the main effect of data set was significant such that /u/ F2 frequency was highest for the Eng-Eng data set, intermediate for the Eng-Span data set, and lowest for the Span-Span data set (all two-way comparisons were highly significant.) This pattern demonstrates that the English vowel space of Spanish-English bilinguals differs in subtle ways from its monolingual counterpart in a direction that reflects the influence of Spanish. Taken together, these findings provide insight into the structure of the mental representations that underlie linguistic sound structure. First, the maintenance of contextual variability across plain and clear speaking styles suggests that low-level signal variability is a systematic feature of the phonetic representations that underlie phonological structure, and that perceptual mechanisms are optimized to extract information that is distributed over time rather than organized into discrete units. Second, the cross-linguistic clear speech tendency toward an expanded vowel space indicates that acoustic distinctiveness can, to some extent, be defined in absolute terms without reference to the size of the segment inventory in question. Finally, the subtle differences between the English vowel spaces of bilinguals versus monolinguals suggests that the two linguistic systems are represented in a single, unified mental model. All of these features point to a model in which phonetic representations encode instance-specific details, and perceptual mechanisms rely on lawful variability in the decoding of the speech signal.

The search for primitives in phonology and the explanation of sound patterns: The contribution of fieldwork studies

Didier Demolin (Free University Brussels)

The search for an adequate set of primitives in phonology and the explanation of sound patterns is a major issue for the discipline. This paper will discuss the contribution of fieldwork studies to such a goal by examining some problems arising from fieldwork studies.

The definition of fieldwork can broadly be understood from two different viewpoints. The broader perspective consists in considering that fieldwork is basically data collecting. The narrower perspective consists in considering fieldwork as collecting data in poorly described or undescribed languages. This paper presents data taken in the narrow sense but without claiming that fieldwork must be restricted to this view. Whatever the viewpoint adopted, the result of such a work should provide reliable data and observations, by using adequate tools, in order to make testable hypotheses through new evidence and natural experiments.

This paper considers that phonological theory is not given 'a priori' but that the underlying primitives and principles must be found by careful observation and by making hypotheses tested through adequate experiments. This view considers, as stated by Lindblom (2000), that the priority of form over substance must be rejected and another paradigm must be found. One consequence is that phonology can be considered as an emergent phenomenon (with phonemic and featural coding being an emergent consequence of optimizing discriminability within a bounded articulatory space) reflecting physical or biologically constraints and evolutionary factors both in ontogeny and in phylogeny. Theoretical phonology must explain how sounds are made, how they change and evolve (i.e. explain the dynamics of sound patterns). Therefore, as stated by Ohala (1990), whatever the model of phonology adopted, phonological theory must be based on models that incorporate parameters coming from the sub-systems involved in speech communication. These are principles relating vocal tract shape and acoustic output, some known aerodynamic principles, and finally some of the principles that explain how our auditory system extracts information from the acoustic signal. In addition, information concerning feedback and control processes should be included in such a theoretical framework. Therefore phonological theory must also incorporate known and well established facts from speech production and speech perception models.

Aerodynamic principles examined come from ejective fricatives (singleton and geminate) in Amharic and from bilabial trills in Central Sudanic languages. Ejectives in Amharic show the interplay between aerodynamic requirements, articulatory control and coordination and the acoustic output. Bilabial trills show that although these sounds might be considered as marginal, the description and explanation of their behaviour require the understanding of aerodynamic, articulatory and acoustic parameters relying on basic principles of speech production. Principles relating vocal tract shape and acoustic output will consider the principle of maximum dispersion in light of the Nuer vowel system, a language exhibiting a set of modal and breathy vowels in its phonemic inventory. The second point will also be illustrated by showing how a model of acoustic/articulatory inversion can shed light on how to find place of articulation and articulatory trajectories. Principles of auditory processing will be used to analyze data from the so-called Lendu vowelless syllables in order to illustrate how the overlap of articulatory gestures can be interpreted perceptually. A case of sound change in Khoisan languages (based on Traill and Vossen 1997) will be examined to show the acoustic and articulatory factors underlying the processes of click replacement and click loss. Finally, cognition and symbolisation in phonology will briefly be examined through a study of word games in Hendo (a Bantu language from Congo). The word games in this unwritten language show manipulations of syllables, phonemes, tones and features. The examination of the learning of these word games show how phonological awareness emerges via a symbolic process. These observations and experiments illustrate how fieldwork studies, if they are understood as more than simple descriptions of exotic sounds, contribute to the refinement for the search of primitives in phonology and to the construction of better models of speech production and perception.

References

- Lindblom, B. 2000. The interplay between phonetic emergents and the evolutionary adaptations of sound patterns. Developmental origins of adult phonology.
- Ohala, J.J. 1990. The segment : primitive or derived ? G. J. Docherty, D. R. Ladd, (eds.) Papers in laboratory phonology II : Gesture, segment prosody. Cambridge University Press. 166-183.

- Traill, A. and R. Vossen. 1997. Sound change in Khoisan languages : new data on click loss and click replacement. Journal of African Languages and Linguistics 18, 21-56.

Acoustic correlates of rhythm classes

Esther Grabe (University of Cambridge) and Ee Ling Low (Nanyang Technological University, Singapore)

Linguists have proposed several classification systems for rhythmic patterns in the world's languages. A dichotomous system has been advocated by e.g. Pike (1945) and Abercrombie (1967). Motivated by a tendency towards isochrony of rhythmic events in languages, these authors distinguish between stress-timed and syllable-timed languages. Dauer (1987) proposed that some languages are more stress-timed, and others are more syllable-timed, and still others are intermediate. Experimental data supporting these hypotheses, however, have been slow to emerge. Moreover, there appear to be rhythmically 'mixed' languages, which exhibit some properties associated with stress-timed languages and some associated with syllable-timed languages. Neither the dichotomous nor the continuous classification system can account for the rhythmic characteristics of such languages. Polish is described as stress-timed, but does not have vowel reduction, a characteristic associated with many stress-timed languages; and Catalan is classified as syllable-timed, but has vowel reduction, cf. Nespor, 1990.

Recently, a set of acoustic measures have been suggested that capture at least some aspects of the rhythmic classifications suggested by linguists. Although these measures do not differentiate rhythmically mixed languages, they show that linguistic classifications of rhythm can be related to acoustic properties of the speech signal. In the present study, we review these measures. On the basis of the evidence available, we propose a new classification system for rhythm. Our system not only distinguishes the rhythmic properties of languages traditionally classified as stress- or syllable-timed, but also includes those which pose a problem for traditional rhythmic classifications.

Our classification system is based on the Pairwise Variability Index (PVI), a measure of variability in vowel duration developed in work by Low and Grabe and colleagues (Low & Grabe, 1995, Low, 1998, Grabe, Post & Watson, 1999, Grabe, Gut, Watson and Post, 1999, and Low, Grabe & Nolan, to appear). Low et al. (to appear) show that the PVI is superior to the standard deviation as an indicator of rhythmicity because the measure is cumulative, controls for changes in speaking rate within intonation phrases, and contains an articulation rate normalisation component. The PVI has been shown to reflect impressionistic comments about a tendency towards syllable timing in Singapore English and French, and a tendency towards stress-timing in German and in British English. In the present study, we have combined the PVI with a measure of rhythmicity suggested by Ramus et al. (1999). Ramus and colleagues showed that the standard deviation of intervocalic intervals can also be related to the syllable-timing/stress-timing dichotomy. The measure we propose involves the combination of the variability of the vocalic intervals in an utterance with that in intervocalic intervals.

The combined PVI has two advantages: Firstly, the measure not only captures the rhythmic characteristics of languages traditionally classified as stress- or syllable-timed but also the characteristics of languages which are rhythmically mixed. Secondly, unlike the dichotomous and continuous classification systems, which require reference to linguistic parameters such as the location of stress or the location of syllable boundaries, the combined PVI is based solely on acoustic evidence.

Figure 1 illustrates the predicted representations for the combined PVI for British English, Singapore English, Spanish, Catalan and Polish (the values shown are hypothetical). Languages such as British English, which exhibit relatively high variability in vocalic and intervocalic intervals, are prototypical for Type HvHi (high vocalic and high intervocalic variability). These are the languages which have been classified as stress-timed. A prototypical LvLi language is Spanish which has been classified as syllable-timed. Compared to English, Spanish exhibits low variability in vocalic and intervocalic intervals, and consequently, timing patterns are more regular. Rhythmically mixed languages such as Polish can be accounted for as Type LvHi. Polish exhibits relatively regular timing in vocalic intervals, but irregular timing in intervocalic intervals. Type HvLi languages might be exemplified by languages such as Japanese (cf. Ramus et al. 1999). Results reported in Low et al. (to appear) support the predictions for British English and Singapore English in Figure 1. The PVI values for British English and Singapore English illustrated in Figure 2 shows that the difference between the two varieties involves a difference in the variability of vocalic. The variability in intervocalic intervals is very small. These data reflect impressionistic comments in the literature about the rhythm of Singapore English.

Our proposal opens up several avenues for research. Firstly, we need to test the predictions of the combined PVI against acoustic data from a wide variety of languages. Secondly, our proposal has implications for work on language acquisition. Mehler et al. (1996) propose that infant speech perception is focused on vowels. Our proposal does not contradict this view; rather, we suggest that infants' perception or rhythm focuses on the variability in (1)

successive vocalic intervals, and (2) the intervals between the vocalic sections. Work by Grabe et al. (1999, to appear) has shown that the rhythm of French, traditionally classified as syllable-timed, is acquired earlier than the rhythm of English, classified as stress-timed. This finding is predicted by the combined PVI: French is an example of an LvLi language and has low variability in vocalic and intervocalic intervals. Such a system is highly predictable and relatively easy to acquire. English, on the other hand, is an HvHi language, characterised by high variability in both parameters. Consequently, rhythmic patterns are difficult to acquire. A prediction which arises is that languages of type HvLi and LvHi should be intermediate in the acquisition process. Thirdly, we need to establish the relative perceptual salience of vocalic and intervocalic variability. Speech perception might be focused equally on vowel durations, as well as on intervocalic intervals, or it might give more weight to the former than the latter.

Figures 1-2

References

- Grabe, E., Gut, U., Post, B., and Watson, I. (to appear). The Acquisition of Rhythm in English, French and German. Proceedings of the Child Language Seminar 1999, London, City University, August 1999.
- Grabe, E., Post, B., and Watson, I. (1999). The acquisition of rhythm in English and French. *Proceedings of the Intonational Congress of Phonetic Sciences*, San Francisco, August 1999.
- Low, E.L. (1998). Prosodic prominence in Singapore English. Unpublished Ph.D. Thesis, University of Cambridge.
- Low, E.L. and Grabe, E. (1995). Prosodic patterns in Singapore English. *Proceedings of the Intonational Congress of Phonetic Sciences*, Stockholm 13-19 August: Vol. 3: 636-639.
- Low, E.L., Grabe, E. and Nolan, F. (accepted). Rhythm in Singapore English and British English. *Language and Speech*.
- Nespor, M. (1990). On the rhythm parameter in phonology. In I. Roca (ed.), *Logical issues in language acquisition*, 157-175, Dordrecht: Foris.

From pitch accent to stress accent in Basque and the typology of accentual systems

José I. Hualde (University of Illinois, Urbana) Gorka Elordieta, Gaminde Inaki (University of the Basque Country), and Raika Smiljanic' (University of Illinois, Urbana)

In this study we first establish that Northern Bizkaian Basque (NB) possesses an accentual system with the following characteristics:

1. There is a lexical contrast between accented and unaccented words, most words belonging to the unaccented class. Lexically accented words are characterized by a fixed accent on a given syllable. Unaccented words do not have prominence on any syllable, except that in isolation or in focus position the final syllable of phrases entirely composed of lexically unaccented words receives accentual prominence.

2. In all phrases, the pitch invariably rises on the second syllable (%LH-) and remains high up to the accented syllable, immediately after which there is a steep fall (H*L). The accented syllable may have the same or a slightly higher pitch than preceding syllables.

3. Accents cause downstepping of following accents, whereas there is no downstep after an unaccented word.

4. Pitch appears to be the only correlate of accent. In particular, there is no durational difference between accented and unaccented syllables.

These properties are established on the basis of the analysis of 150 utterances x two speakers and confirmed with smaller samples from 10 more speakers. Our results are also consistent with previous descriptive work (Hualde 1991, Jun & Elordieta 1997) and can now be taken as proven facts about the NB accentual system. This prosodic system presents striking similarities with Tokyo Japanese (Beckman & Pierrehumbert 1986).

Other western and central Basque dialects differ considerably in their prosodic properties from the description above. In fact, some of these other dialects are prosodically much more similar to Castilian Spanish (in which all speakers are bilingual). There is solid evidence, however, that all accentual systems in the western and central Basque area are historically related and that, in fact, NB is the most conservative of them (see Hualde 1990, 1995). In the second part of this paper, we report on an instrumental study of a number of local dialects situated between the two extreme cases. What we find is that dialects differ very gradually in the extent to which they preserve the properties listed above for NB.

One feature in which NB differs markedly from Spanish is in presenting a non-accentual rise on the second syllable and a high plateau up to the accented syllable. This contour has been reinterpreted in two different ways. In some areas the non-accentual rise on the second has been eliminated. Two stages in this evolution are apparent: In some dialects (Larrabetzu, older Arratia speakers) the pitch rises gradually from the beginning to the accented syllables. In a more advanced state in the evolution (younger Arratia speakers, Antzuola), the pitch remains low up to the onset of the accented syllable, as in Spanish.

A different evolution has been the reinterpretation of the non-accentual rise on the second syllable as signaling the position of the accented syllable. This has resulted in systems with accent on the second syllable in all or most words.

By analyzing F0 contours in a sets of examples which in NB minimally contrast in the presence vs. absence of a lexical accent (e.g. lagunen alab'ak 'the friend's (sg) daughters' (1 accent) vs. lag'unen al'abak 'the friends' (pl) daughters' (2 accents)) we show that an independent phenomenon in some non-NB varieties is the loss of the lexical accented/unaccented distinction (all major class words becoming accentable, as in Spanish).

We draw two major conclusions for the typology of accentual systems. First of all, the striking coincidence between NB and Tokyo Japanese in a number of important prosodic properties suggests that this set of common properties can be used to characterized an accentual prototype ("pitch-accent language"). On the other hand, the existence in theBasque territory of varieties presenting a range of intermediate points between NB and Spanish shows that it is not fruitful to impose a clear-cut distinction between stress-accent languages and pitch-accent languages. Rather, a typology in terms of focal types agrees better with the facts.

Finally, as a point of comparison for our view on the typology of accentual systems, we briefly consider some

aspects of dialectal variation in Serbo-Croatian. Although sometimes referred to as a pitch-accent language (which would place it in the same prosodic type as Japanese and NB Basque), Serbo-Croatian is better classified as a stress-accent language with a lexical contrast in the alignment or type of intonational tonal contours associated with stressed syllables (pitch-accents in the terminology of Pierrehumbert 1980) restricted to words stressed on the initial syllable. Swedish/Norwegian and Dutch/German dialects would fall in the same class of accentual systems. Some Serbo-Croatian dialects have lost this pitch-accentual contrast. We show that in languages of this type the loss of the accentual contrast results in a more typical stress-accent system.

Predicting the relative importance of prosodic cues from linguistic structure: The case of stress and tone in the word prosody of Samate Ma`ya

Bert Remijsen (Leiden University)

Lexical stress can be marked by a variety of prosodic parameters, such as duration, fundamental frequency (F0), vowel quality and spectral balance. According to the Functional Load Hypothesis (Berinstein 1979), the importance of specific prosodic parameters is not a fixed linguistic universal, but instead depends language specifically on phonological structure. In a tone language, for instance, fundamental frequency (F0) cannot be a stress correlate, because it is already used to distinguish the paradigmatic tonal contrast. In the strongest form of the Functional Load Hypothesis (FLH), the potential use of a prosodic parameter as a stress cue is precluded if it encodes other phonological distinctions in the sound system of the language. On the basis of a study of the acoustic encoding of stress in the tone language Thai, Potisuk, Gandour and Harper (1996) have argued that this formulation of the FLH is too strong, and that, instead, it should be interpreted in a more relative fashion. They found duration to be the main stress correlate, although Thai also has a small set of word pairs with a long-short vowel contrast. They conclude that "[t]he question is not simply whether an acoustic parameter is implicated in lexical contrasts, but rather what the degree of its involvement is at the lexical or morphological level." [Potisuk et al. 1996:211]. Though this conclusion is appealing, the acoustic evidence is flawed: the contrast Potisuk et al. investigated is not a lexical minimal pair, but one between (i) two monosyllabic phonological words and (ii) a compound derived from the same two words. Consequently, the observed variation in duration, the cornerstone of Potisuk et al.'s argument, can also be attributed to anticipatory shortening and/or final lengthening. In order to test the validity of the FLH in a tone language, we have to study a tone language with lexically rather than phrasally contrastive stress. Claims that such languages exist have been made -- e.g. for Papiamentu (Römer 1991), but have never been supported by acoustic evidence.

I claim that the Austronesian tone language Ma`ya (Samate dialect) features lexically contrastive stress. This claim is based on (i) a study of the word prosodic phonology and (ii) an acoustic investigation of minimal stress pairs. This rare typological structure makes Ma`ya an ideal test-case for the FLH.

Alternative analysis of the phonology The phonology of Samate Ma`ya has been described before by van der Leeden (1993). He analyzes Ma`ya as a four-tone language with stress dependent on tone. Van der Leeden's study contains a number of complexities, which can be explained more elegantly by assuming that the language features contrastive lexical stress. By assuming that Ma`ya has contrastive stress, (i) one can analyze the language as a three-tone instead of a four-tone system, (ii) the word-prosodic pattern of loan words no longer requires a stress shift rule, and (iii) exceptions to a tonal reduction process become regular.

Acoustic investigation Extended Swadesh lists were recorded from four native speakers of Samate Ma`ya. An acoustic study of the vowels of three minimal stress pairs [example 1] in these lists shows that stress is marked predominantly by duration, which is precisely the prosodic parameter the FLH (in either the strong or the relative version) predicts to be the strongest stress correlate in a tone language. Statistical analysis (repeated measures ANOVA and LDA) reveal that the lexical stress contrast is encoded by duration, spectral balance and vowel quality, in that order of importance [fig.1A]. F0, on the other hand, which is employed to encode paradigmatic tonal contrast, does not vary systematically in function of stress. Precisely the reversed rank order is found for the cueing of a -- three-member -- lexical tone contrast [fig.1B].

In summary, both the phonological analysis and the results of the acoustic study support the claim that Ma`ya features contrastive lexical stress in addition to tone. As such, it is the first language for which such a claim is experimentally supported. On the basis of the relative importance of the prosodic parameters in encoding stress and tone, I present a more specific interpretation of the FLH.

Example 1

Penultimate stress	Final stress
`tala3	*ta`la3*
'banana'	'k.o. plant'
`kaya3	*ka`ya3*

'rich' (Malay loan word)	'machete'
`mana3	*ma`na3*
'light (of weight)'	'oil, grease'

Note on transcription: Lexical stress is marked by a ` preceding the stressed syllable. The tone is transcribed after the syllable in a numeric notation; 3 refers to a high tone.

Figure 1

References

- Berinstein, A.E. (1979). A Cross-Linguistic Study on the Perception and Production of Stress. UCLA Working Papers in Phonetics 47.
- Leeden, A.C. van der (1993). Ma`ya: A Language Study: A. Phonology. Jakarta - LIPI-RUL Series 2A.
- Potisuk, S., Gandour, J., and Harper, M.P. (1996). Acoustic Correlates of Stress in Thai. Phonetica Vol. 53, 200-220.
- Römer, R. (1991). (eds. N.H. Smith and J.M. Stewart) Studies in Papiamentu tonology. Caribbean Culture Studies 5. Amsterdam and Kingston - Amsterdam.

Underspecified recognition

Aditi Lahiri (University of Konstanz)

The speech signal of the same phonetic segment varies across dialects and speakers, within speakers between segmental and prosodic contexts, and even for the same speaker and context with repetition, speaking rate, emotional state, microphone and line condition, etc. Any theory of lexical phonological representation and processing must be able to account for productive phonological processes such as assimilation and deletions, particularly across word boundaries. Explicitly or implicitly, all such theories assume that at the level of the lexical entry there is a single abstract representation, so that not every phonological surface variant form is listed. This leaves unanswered, however, the question of precisely how the system does recognize the different phonetic variants of the word when the relationship between these realisations and the lexical entry is not transparent. We will consider here the linguistic, psychological, and computational adequacy of two very different approaches to answering this question.

The differences in these two approaches are the consequence of the different claims they make about the nature of the representation, choosing between fully and partially specified representations. If the representation is fully specified and matches the 'ideal' (or citation) realisation of the word, then this requires the postulation of additional mechanisms to explain how the other variants of the word are dealt with. For instance, if the word SWEET is represented with a fully specified [t], the matching of [swik] in the phrasal context "sweet girl" would have to be resolved by other mechanisms. If the representation is, in contrast, underspecified, so that it can incorporate all variants of the word, then this raises questions about how the system makes choices between different variants and how lexical ambiguity is resolved. Thus, if the representation of /t/ has no place information specified, the phonetic form [swip] would match both lexical /swip/ and /swit/. Depending on the representational choice, the further issues concern the intermediate representation (if any) and the mechanisms involved in identifying the correct lexical items. For the fully specified approach, rules of inferencing would be necessary pre-lexically, so that all possible lexical candidates are included. Thus, if the system encounters a [k], a right context rule would infer that the lexical candidates could either end with a /k/ or a /t/ - the choice would be delayed till the right context is determined. For the underspecified approach, there need be no rule of inferencing; rather the representation itself (along with an universal definition of matching) would allow both candidates to be activated.

The adequacy of the two types of approach will be critically evaluated, for a wide range of productive phonological processes, looking at their linguistic motivation, at their computational tractability, and at their success in predicting psycholinguistic performance. A fully specified rule-based approach would make different predictions. In evaluating the two hypotheses, we draw on evidence from psycholinguistic experiments and language change. We will present our model -FUL (Featurally Underspecified Lexicon) - developed together with Henning Reetz, which is also implemented in a recognition system assuming an underspecified lexicon and a ternary matching procedure.

Phonetic features in ASR: A linguistic solution to acoustic variation?

Jacques Koreman, Bistra Andreeva and William Barry (University of the Saarland)

In most phonological theories, phonemes are considered as a set (or hierarchy) of (possibly underspecified) phonetic features, which are the minimal number of formal properties needed to distinguish the phonemes in the language system from each other.

In most state-of-the-art automatic speech recognition (ASR) systems, however, phonetic features do not play any role. The statistical models for each phone or phoneme are based on a spectral parameterisation of the signals, like mel-frequency cepstral coefficients (MFCC's) and energy.

Three questions are dealt with in this paper:

1. Can we successfully bridge this gap between phonological theory and ASR by using phonetic features in ASR?
2. Which phonetic feature set is most appropriate for ASR?
3. Can we attain the same result by using more complex non-linguistic modelling?

1. PHONETIC FEATURES IN ASR

To bridge the gap between phonologists' formal representation of the phoneme and the almost purely acoustic description of the signal used in ASR systems, we have used phonetic features to create statistical phone models for automatic speech recognition. The phonetic features were derived from the spectral representation of the signal used in most standard ASR systems (MFCC's + energy) by means of a neural network.

Not only do we find a clear increase in the phoneme identification rate (see under 2 below) [1], the confusions between phonemes are also much easier to interpret, since phonemes which are confused are usually very similar in terms of the phonetic features they are made up of. This is not the case when acoustic parameters are used to create phoneme models [2].

2. DIFFERENT PHONETIC FEATURE SETS

It is not self-evident which set of phonetic features is most appropriate to describe phonological categories and the processes that operate on them, since the various feature theories have different phonological implications. To evaluate how appropriate the different feature sets are for application in an ASR system, we have used several different feature sets, both articulatory-phonetic (IPA) and phonological (SPE) [3]. We have so far compared the phoneme identification results for both underspecified and fully specified SPE features with those for the set of features used in the IPA to distinguish all phonemes. In addition, the results were compared to those in a standard ASR system using acoustic parameters (MFCC's) directly to create phone models. We found a clear improvement in the phoneme identification rate when phonetic features were used to model the phones, in comparison to directly using acoustic parameters. Underspecified SPE features led to the best performance (for multi-language Eurom0 data, without the use of a lexicon or language model) of all:

acoustic parameters:	15.6%
IPA features:	42.6%
SPE features:	36.2%
Underspecified SPE features:	46.1%

In addition to the features sets reported so far, the phoneme identification results for articulatory features [4] will be reported and their relative merits will be discussed.

3. VARIATION MODELLING VERSUS LINGUISTIC MODELLING

The acoustic-phonetic mapping in a neural network combines two advantages, namely

1) variation modelling: different acoustic realisations of the same phoneme (e.g. allophonic variants) can be

discerned by the neural network

2) linguistic modelling: these different realisations are mapped onto more homogeneous, distinctive features

Even if the neural network can reduce the variation in the input parameters for statistical modelling by mapping different acoustic realisations of a phoneme onto phonetic features, the question remains whether the same result can be reached by using a non-linguistic approach. Variation modelling can also be achieved by using more complex acoustic phoneme models (multiple mixtures per state in HMM), so that we do not necessarily have to make a mapping onto phonetic features to achieve this goal. A comparison of the performance of a standard system which does not use phonetic features with the performance of a system in which phonetic features are used to train the phoneme models shows the merits of using a signal representation derived from phonological theory.

References

1. Koreman, J., Barry, W.J. & Andreeva, B. (1998). "Exploiting transitions and focussing on linguistic properties for ASR," Proc. Int. Conf. on Spoken Language Processing (ICSLP'98), Sydney.
2. Koreman, J., Andreeva, B. & Barry, W.J. (1998). "Do phonetic features help to improve consonant identification in ASR?" Proc. Int. Conf. on Spoken Language Processing (ICSLP'98), Sydney.
3. Koreman, J., Andreeva, B. & Strik, H. (1999). "Acoustic parameters versus phonetic features in ASR," Proc. of the 14th Int. Congress of Phonetic Sciences (ICPhS'99), San Francisco.
4. Deng, L. & Sun, D. (1994). "A statistical approach to automatic speech recognition using the atomic speech units constructed from overlapping articulatory features," J. Acoust. Soc. Am. 95(5), 2702-2719.

Explosives, implosives, and nonplosives: Some linguistic effects of air pressure differences in stops

Nick Clements (CNRS, Paris) and Sylvester Osu (LLACAN-CNRS, Villejuif)

This paper concerns the phonological and phonetic analysis of the class of non-explosive stops. This class notably includes what are traditionally called implosives, (pre)glottalized stops, laryngealized stops, lenis stops, and others. However, these sounds are still not well understood. Unanswered questions include the following: How many phonetic types of such sounds can be distinguished? How many of these types are phonologically contrastive? How do these sounds pattern phonologically? What is their feature analysis? New evidence bearing on these questions comes from an ongoing study of Ikwere [ìkwéré], an Igboid language spoken in the Rivers State in Southeast Nigeria currently being conducted by the two authors. Ikwere has two pairs of voiced and voiceless bilabial stops, written p, b, kp, gb in the current orthography. The sounds kp, gb are reflexes of older labiovelar stops reconstructed for Proto-Igboid by Williamson (in press). However, in the dialect studied here, they are realized as bilabial sounds with no velar contact at any point in their production. How, then, are they distinguished phonetically and phonologically from the "ordinary" bilabials p and b?

To answer this question we have carried out an acoustic, aerodynamic and articulatory study of these sounds. This study shows these sounds to have a number of properties lacking in their "ordinary" bilabial counterparts: (i) they are produced with little or no increase in intraoral air pressure during closure, (ii) the voiceless sound kp is produced with tight glottal closure, is often imploded at release, and always shows a brief period of prevoicing preceding release, (iii) the voiced sound gb is fully voiced throughout and tends to show an increase in voicing amplitude during closure; (iv) both sounds are labialized (or labiovelarized), especially at release; (v) articulation is lax, and (vi) they have partly (kp) or fully (gb) nasalized realizations before phonemically nasal vowels. In some respects, these sounds resemble lax or lenis stops; however, they are not consistently shorter than their explosive counterparts as the feature lax requires (Jessen 1998). In others, they resemble the voiced and voiceless implosives of languages like Lendu and Owere Igbo; however, unlike true implosives, they are not produced with a glottalic airstream mechanism, and while kp is sometimes imploded at release, gb never is in our data.

On the basis of these characteristics it is concluded, following an earlier suggestion by Stewart (1989) regarding the lenis stops of Ebrié, that the sounds in question constitute a class of nonobstruent stops. Obstruents are defined, following Halle (1992) and others, as sounds produced with an increase in oral air pressure during their production. Nonobstruent stops such as true implosives and Ikwere kp, gb are characterized by little or no increase in oral air pressure during their production and the absence of a plosive burst at their release (hence their characterization as "nonexplosive stops"). Most other phonetic and phonological properties of these sounds (such as their failure to induce tone depression, or the fact that they frequently alternate with nasals and liquids) follow directly from their nonobstruent status.

It is proposed that true implosives, (pre)glottalized stops, and laryngealized stops, as well as "lenis" stops (as described for Ebrié) and nonexplosive stops (as in Ikwere) may constitute variant phonetic implementations of a single phonological class of nonobstruent stops. Within this class, the commonest phonological contrast at any single place of articulation is that between glottalized and nonglottalized sounds. Whether any further phonological distinctions need to be made within this class remains an open question.

References

- Halle, M. 1992. "Features," in W. Bright (ed.), Oxford International Encyclopedia of Linguistics, vol. 3. New York: Oxford University Press, pp. 207-212.
- Jessen, M. 1998. Phonetics and Phonology of Tense and Lax Obstruents in German. Amsterdam: John Benjamins.
- Stewart, J.M. 1989. "Kwa." In Bendor-Samuel (ed.), The Niger-Congo Languages, New York: Lanham, pp. 217-45.
- Williamson, K. in press. "Reconstructing Proto-Igboid Obstruents." In V. Carstens and F. Parkinson, eds., Trends in African Linguistics 4 (Proceedings of ACAL 28). Trenton, N.J.: Africa World Press.

The phonetics of phonologization

John Ohala (University of California, Berkeley)

The most important stage in the development of phonetically natural sound patterns is the process called by Jakobson 'phonologization', i.e., where purely phonetically and contextually predictable phonetic variation becomes non-predictable and an independent part of the pronunciation either by its presence or its magnitude. This occurs, I maintain, due to a misparsing or misinterpretation of the speech signal by the listener: the phonetically predictable link between phonetic cause and phonetic effect is not recognized and the listener thus forms a novel lexical representation (which, probabilistically, could spread to other speakers, other words, etc. via social and psychological means). However, in some cases it may be difficult to differentiate between a case where the link between a phonetic cause and its effect is phonetic, i.e., due to physical factors, and where the link is phonological, i.e., due simply to the co-occurrence of the two phonological entities. In this paper I outline a general method for making this differentiation and review several prior studies that exemplify this method.

Assimilatory processes and aerodynamic factors

Maria Josep Solé (Autonomous University of Barcelona)

In the search for the phonetic basis of assimilatory processes, a number of factors have been investigated, including articulatory (Browman and Goldstein 1991, Recasens et al. 1997), auditory (Jun 1995, Hume et al. 1999) and perceptual properties (Ohala 1990). The work presented here explores the role of aerodynamic factors in some assimilatory processes. Specifically, it investigates sequences of fricatives and trills produced with the same and different articulators, in order to account for the common assimilation of lingual fricatives to following apical trills (e.g., /z, s, zh, sh/ + /r/ > /r:/), where the fricative disappears in a variety of languages. Consonants other than fricatives accommodate to the constriction location of the trill but remain stable otherwise. This study tests the hypothesis that the extensive anticipatory movements for the trill override the critical articulatory positioning and time required for the acoustic event of frication.

Simultaneous EPG, EGG and acoustic data were obtained for three Catalan speakers producing a variety of word final lingual fricatives followed by trills. Prosodic boundaries were varied to observe the effect of time constraints on fricative production. Across minor boundaries, varying degrees of fricative to trill assimilation, with no resulting frication, reflected differences in the timing of the sequential inputs. Across major boundaries, full fricative realization reflected input strings of the size of the tone group. The results suggest that the early onset of the movements for the trill overrides the time required to achieve the aerodynamic conditions required to generate turbulence for fricatives.

In order to determine the time needed to achieve the pressure drop at the oral constriction for audible frication (T ? Pfrication), in a second experiment oro-pharyngeal pressure and airflow were recorded simultaneously in two subjects producing steady state and intervocalic fricatives and trills. The time from onset of Po rise (i.e., onset of articulatory movements for the fricative) to onset of audible frication was measured. Under normal conditions, T ? Pfrication was in the range of 50 ms for voiced and 33 ms for voiceless fricatives. Voiced fricatives take longer to achieve a sufficient pressure drop at the constriction, vis-à-vis voiceless fricatives, due to glottal impedance and reduced transglottal flow. Hence, if onset of articulatory movements for the trill reach the articulator within 50 ms from onset of the movements for the voiced fricative, audible friction will not be achieved. These predictions are in accord with the results observed: the early onset of movements for the trill bleed the time required to generate frication.

The assimilation of fricatives to following trills can thus be modelled as a duration dependent effect along the lines of Lindblom's (1963) undershoot model: As the time to build up the oral pressure for frication becomes shorter, due to anticipatory movements for the trill, the pressure drop required for frication is not achieved. Hence, both aerodynamic and acoustic lack of turbulence.

In summary, the data indicate that the acoustic requirements (audible turbulence) for fricatives impose relatively strict aerodynamic and time constraints (sufficient rate of flow and duration of constriction) which may be adversely affected by competing overlapping gestures, resulting in lack of frication. Thus assimilatory models should include degree of aerodynamic constraint and temporal constraints.

The implications for coarticulatory theories, assimilatory models and historical sound change are considered.

References

- Browman, C.P. and Goldstein, L. 1990. Tiers in articulatory phonology, with some implications for casual speech. In D.R. Ladd and G. Docherty (eds.), Papers in Laboratory Phonology II, 341-376. Cambridge: CUP.
- Jun, J. 1995. Place assimilation as the result of conflicting perceptual and articulatory constraints. Proceedings of WCCFL 14, 221-237.
- Hume, E., Johnson, K., Tserdanelis, G. & Seo, M. 1999. Proceedings of the XIV International Congress of Phonetic Sicences. San Francisco, Vol 3, 2069-2072.
- Lindblom, B. 1963. A spectrographic study of vowel reduction. Journal of the Acoustical Society of America, Vol. 35. 11, 1773-1781.
- Ohala, J.J. 1990. The phonetics and phonology of aspects of assimilation. In J. Kingston & M.E. Beckman (eds.), Papers in Laboratory Phonology I, 258-275. Cambridge: CUP.

- Recasens, D., Pallarès, M.D. and Fontdevila, J. 1997. A model of lingual coarticulation based on articulatory constraints. Journal of the Acoustical Society of America 102.1, 544-561.

Gestural overlap and recoverability: Articulatory evidence from Georgian

Ioana Chitoran (Dartmouth College), Louis Goldstein (Yale University and Haskins Laboratories), and Dani Byrd (University of Southern California)

Recent investigations of gestural patterning (Byrd 1996) have found that consonant gestures exhibit less temporal overlap in a syllable/word onset than in a coda or across syllables. One possible account for this difference is that substantial overlap of obstruent gestures may threaten their perceptual recoverability, particularly in utterance-initial position, where no acoustic information is provided during the formation of the consonant gestures (no VC transitions). Additionally, recoverability considerations may account for results of Byrd (1992, 1996), Zsiga (1994), Surprenant and Goldstein (1998) who show that front-to-back order of place of articulation in stop-stop sequences (labial-coronal, coronal-dorsal, labial-dorsal) allows more overlap than the opposite order. Presumably, the C1 release would produce no acoustic manifestation if C2 were already formed just in case C2 constricts the vocal tract anterior to C1. If C2 were posterior to C1, then at least some acoustic information would be generated on release of C1 (even without a substantial release burst). Similar recoverability requirements are proposed here to account for consonant sequencing phenomena in Georgian, which violates the sonority sequencing generalization.

Previous articulatory investigations of gestural patterning were undertaken in English, therefore the data on the effect of position are limited. The only word-initial sequences are /s/-stop clusters. They have special phonological properties, and are sometimes argued to constitute single segments (Ewen 1982). This "special" status, rather than recoverability considerations, could account for the differences in overlap. We therefore tested for the positional effect in Georgian, which allows a variety of stop-stop sequences in onsets, and elsewhere. We also tested the hypothesis that front-to-back ordering of stops allows more overlap than back-to-front ordering. This hypothesis has previously been tested in an acoustic study in Georgian (Chitoran 1999) by measuring the inter-burst interval between C1 and C2. The interval was found to be significantly shorter in front-to-back than in back-to-front sequences, suggesting a higher degree of overlap in the former. However, acoustic data do not provide an unambiguous measure of gestural overlap.

This paper reports on a magnetometer study of gestural overlap in Georgian. C1C2 sequences were examined as a function of position in the word, and of the order of place of articulation of C1 and C2. Two native speakers of Georgian served as subjects. Example stimuli are shown below.

(1)	word-initial	word-internal	across words
a. front-to-back			
	ptila	aptari	
	dgeba	adgeba	albat#kari
	bgera	abga	
b. back-to-front			
	t'ba	dat'borva	albat#bani
	gdeba	agdeba	
	gberavs	dagbera	

The recoverability-based predictions were borne out. Effects of both position in the word and of order of place of articulation were obtained. A significant difference was found in the timing of the onset of the C2 gesture relative to C1. C2 was found to overlap a greater proportion of C1's movement duration word-internally and across words than word-initially. As predicted, more overlap was allowed in positions where recoverability of C1 is less easily compromised.

More overlap was also found in sequences with a front-to-back order of articulation (1a) than in back-to-front sequences (1b), therefore where C1 gestures are less easily hidden by C2 gestures. These results may also explain the impressionistic descriptions of Georgian so-called "harmonic clusters" as single segments. These are very common front-to-back C1C2 sequences sharing laryngeal features. The large amount of overlap allowed in them may be responsible for the single segment claim, and for the observation that they share laryngeal features. Assuming that only the C1 gesture can have a laryngeal gesture coordinated with it, that single laryngeal gesture may still be active during C2, if C1 and C2 overlap substantially. However, if C1 and C2 are minimally overlapping, then the C1 laryngeal gesture would not be active during C2, and a default laryngeal state should

result. This default state (usually voiceless unaspirated) actually occurs for C2 in non-harmonic clusters.

Georgian syllable onsets violate the sonority sequencing generalization but are apparently sensitive to gestural recoverability requirements as reflected in overlap patterns. Thus, while sonority sequencing has the function of allowing gestures to overlap while still allowing recoverability (Mattingly 1981), this function can apparently be filled in other ways.

References

- Byrd, D. (1992) Perception of assimilation in consonant clusters: A gestural model. Phonetica 49, 1-24.
- Byrd, D. (1996). Influences on articulatory timing in consonant sequences. Journal of Phonetics, 24, 209-244.
- Chitoran, I. (1999) Accounting for sonority violations: the case of Georgian consonant sequencing. Proceedings of ICPhS 99, 101-104.
- Ewen, C.J. (1982). The internal structure of complex segments. H.van der Hulst & N.Smith (Eds.), The structure of phonological representations. Dordrecht:Foris Publications, 27-67.
- Mattingly, I.G. (1981). Phonetic representation and speech synthesis by rule. T.Myers, J.Laver, & J.Anderson (Eds.), The cognitive representation of speech (415-420). Amsterdam:North Holland.
- Surprenant, A.M. and L. Goldstein (1998) The perception of speech gestures. JASA 104, 518-529.
- Zsiga, E.C. (1994) Acoustic evidence for gestural overlap in consonant sequences. Journal of Phonetics 22, 121-140.

Tonal association and target alignment: Implications for intonation theory

Sónia Frota (University of Lisbon)

In European Portuguese declarative intonation, there are two HL nuclear sequences that differ both in their intonational form and meaning, as illustrated in (1). The present work, following proposals made in Frota (1998), discusses the phonological status of the HL patterns, as well as their phonetic interpretation, within the autosegmental-metrical theory of intonation (e.g. Ladd 1996). Further, some implications of the data observed and the analysis proposed for the intonational phonology view will be explored.

The presence of 'different' tonal patterns, one conveying a *news sentence* semantics and the other conveying (free narrow/contrastive) focus seems unmistakable in European Portuguese (EP). However, questions such as how the H and L targets are phonological organised and what are the phonetic exponents of phonological association (or starredness) require examination. Is the difference found best captured in terms of the presence of a prosodic boundary that has a tonal manifestation (particularly in the focus case - cf. Hayes & Lahiri 1991, Ladd 1999) versus a purely accentual tonal manifestation? Or is it a whole accentual matter to which the phonological structure of the pitch accents involved is the key (Grice 1995b, Arvaniti et al. 1998, 2000)?

A bitonal analysis of the HL patterns, grounded on an association distinction, is argued for on the basis of tonal alignment with the segmental string, alignment stability, and slope and timing relations between tonal targets: (i) the consistent phonetic distance between H and L in either sequence irrespective of prosodic word or phonological phrase boundary location, rules out the tonal edge hypothesis; (ii) the fact that one of the targets always precedes/follows closely the target that is aligned with the stressed vowel is a strong indication that both sequences behave as complex pitch accents; (iii) the finding that only in the focus sequence (H* L), but not in the neutral sequence (H L*), the alignment and scaling of the starred tone affects the realisation of the unstarred element supports the hypothesis that the distinction resides in the phonological structure of the two accents.

The bitonal analysis H+L* and H*+L, arguably the only one compatible with the data, has interesting consequences for some of the ongoing debates in intonation theory, such as the pitch accent structure proposals and the status of leading and trailing tones, or the levels versus configurations view of intonation primes. It will be shown that leading and trailing tones cannot be simply seen as the unstarred elements of bitonal accents, as both headedness and structural relations (namely, dominance) between tones do matter: besides either preceding or following the accent's head, the unstarred elements may *lean* on the head or be *independent* of the head, a realisational difference that is adequately captured by a representation of unstarred tones as elements projected *within* or *outside* the accent's head scope. The EP facts thus favour a hierarchical-structured analysis of pitch accents (along the lines of Grice 1995a) over a linear view (as in the standard Pierrehumbertian approach) or a flat-structured proposal (as in Grice 1995b). Further, under this analysis the bitonal character of nuclear accents in Portuguese declaratives fits nicely into an accentual head-dependent asymmetry at tune level (cf. Dresher & van der Hulst 1995). Last but not least, the features of the HL bitonal accents offer a clear argument against the view that configurations, and not tonal targets, are the primitives of intonation: the distinction found is unexpected and unexplained under the configurations view, as it cannot be accounted for by the properties of pitch movement or shape alone.

The contrast between the HL accents of EP lends support to the reanalysis of H+L* and H*+L as 'real' bitonal events (in the sense that both tones have a clear phonetic interpretation as H and L targets), against their problematic status in the original Pierrehumbert's taxonomy as the only bitonal events whose targets are not independently identifiable in the contour (cf. Pierrehumbert 1980, Beckman & Pierrehumbert 1986). In this respect, Portuguese joins other romance languages that, unlike English, have also been reported to make a similar use of bitonal HL accents (cf. Sosa 1991 for American Spanish; Grice 1995a for Palermo Italian; D'Imperio 1996, 1997 for Neapolitan Italian). Any basic intonational taxonomy which aims at a cross-language comparison of intonation should therefore include such bitonal events.

Figure 1

References

- Arvaniti, Amalia, D. Robert Ladd & Ineke Mennen. 1998. Stability of tonal alignment: the case of Greek prenuclear accents. *Journal of Phonetics* 26: 3-25.
- Arvaniti, Amalia, D. Robert Ladd & Ineke Mennen. 2000. What is a starred tone? Evidence from Greek. In M. Broe and J. Pierrehumbert (eds.) *Papers in Laboratory Phonology V*. Cambridge: CUP, 119-131.
- Beckman, Mary and Janet Pierrehumbert. 1986. Intonational Structure in Japanese and English. *Phonology Yearbook* 3: 255-310.
- D'Imperio, Mariapaola. 1996. Caratteristiche di timing degli accenti nucleari in parlato italiano letto. Associazione Italiana di Acustica, XXIV Convegno Nazionale, Trento, 55-60.
- D'Imperio, Mariapaola. 1997. Bredth of focus, modality and prominence perception in Neapolitan Italian. *OSU Working Papers in Linguistics* 50: 19-39.
- Dresher, B. Elan & Harry van der Hulst. 1995. Head-Dependent Asymmetries in Prosodic Phonology. Ms., University of Toronto and HIL/University of Leiden.
- Frota, Sónia. 1998. Prosody and Focus in European Portuguese. PhD dissertation, University of Lisbon.
- Grice, Martine. 1995a. *The intonation of interrogation in Palermo Italian: implications for intonation theory*. Tübingen: Niemeyer.
- Grice, Martine. 1995b. Leading tones and downstep in English. *Phonology* 12: 183-233.
- Hayes, Bruce & Aditi Lahiri. 1991. Bengali Intonational Phonology. *Natural Language and Linguistic Theory* 9: 47-96.
- Ladd, D. Robert. 1996. *Intonational Phonology*. Cambridge: CUP.
- Ladd, D. Robert. 1999. Segmental Anchoring of Tonal Targets: Some Consequences. Conference given at the University of Lisbon.
- Pierrehumbert, Janet. 1980. The phonology and phonetics of English intonation. PhD dissertation, MIT.
- Sosa, Juan M. 1991. Fonética y Fonologia de la entonación del Español Hispanoamericano. PhD dissertation, University of Massachusetts, Amherst.

Abstracts

Temporal distribution of interrogativity markers in Dutch: A perceptual study

Vincent van Heuven (Leiden University) and Judith Haan (University of Nijmegen)

Summary

Our research aims to establish the phonetic correlates of the phonological (and semantic) contrast in sentence type between statement and question. The object language is Dutch. The archetypical question marker is the sentence-final rise H%. However, there is an increasing body of evidence showing that the marking of interrogativity is distributed over the entire course of the spoken utterance such that the percept of question intonation builds up gradually. The phonetic implementation of the question morpheme therefore seems in need of refinement. We argue that a superposition model with global up-/downtrend of pitch and local H-targets on accents and boundaries is best suited for this task.

Earlier work

Haan et al. (1997, 1998) have shown that the prosodic encoding of interrogativity is stronger as the sentence contains fewer lexico-syntactic question markers. Question intonation is strongest in declarative questions (DQ), i.e., questions that are lexico-syntactically identical to statements. Dutch statements (ST) were found to differ from their DQ counterparts in both local and global intonational aspects. In terms of local pitch movements DQ is characterized by a sentence-final pitch rise (H%). Also, whilst the pitch-accents on subject (first noun) and object (last noun) are about equal-sized in ST, the subject accent is considerably smaller (both in excursion size and in peak height) and the object-accent larger than their counterparts in DQ. A global difference is manifest in the presence (in ST) versus absence (in DQ) of downtrend in the L targets (see fig. 1), i.e. in the slope of the 'low declination line' (see also Thorsen, 1980).

Results of a gating experiment using natural utterances (van Heuven et al., 1997; see fig. 2) showed that Dutch listeners clearly distinguish ST from DQ by the time the pitch-accent on the object is audible. This shows that H% is not needed for Dutch listeners to determine (or a least predict) the interrogative mode of a spoken sentence. However, we still do not know what aspect(s) of the utterance allow this early discrimination of ST and DQ: is it (i) the absolute size of the object accent, (ii) the relative difference in accent size between subject and object, or (iii) the (lack of) global downtrend from sentence onset towards the object accent? This specific question is addressed in the present experiment.

Experiment

The present study was set up to establish the relative importance of each of the above acoustic cues, as they develop over the time-course of the utterance, in a 'gating' experiment with resynthesized exemplars of a natural utterance, systematically manipulating (i) size of subject accent (0, 4, 8, 12 semitones), (ii) size of object accent (same values), global down/uptrend of lower reference line (-3.0, -1.5, 0, +1.5 st/s), and presence/absence of final pitch rise (H%=0 or 8 st).

The resulting 128 utterance types were truncated at five points in time: (1) after the subject accent, (2) before the object accent, (3) after the object accent, (4) before H%, (5) at end of sentence. Stimuli were presented in random order but shorter fragments always preceded longer ones. Two groups of listeners (20 for gates 1 through 4, 20 for gate 5) judged for each fragment/sentence whether it was (part of) a question or a statement, with binary forced choice.

Results and conclusion

The results bear out that H% is the strongest interrogativity cue, but that global uptrend and size of the object accent are important secondary cues (see figs. 3-4-5). Size of subject accent (not shown) is only a question cue if it can be interpreted as a contrastive accent (apparently, narrow focus is conducive to hearing a question). ST and DQ are successfully differentiated by the time the accent on the object is heard (gate 3). Crucially, when the pitch peak on the object accent is held constant, global uptrend combined with a small excursion size elicits more question responses than downtrend combined with a large object accent. This shows that global down-/uptrend of pitch is the stronger (and rather indispensable) early interrogativity marker. Our results argue for the inclusion of a global

trend parameter in intonation models (of Dutch), to be used independently of such mechanisms as downstep/upstep and accent scaling (reflecting greater or lesser focus on the constituent bearing the accent).

Acknowledgement

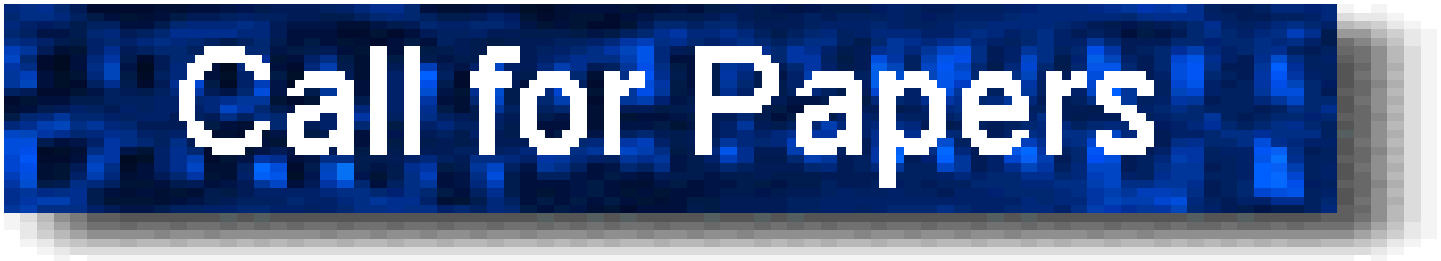
This research was funded in part by the Netherlands Organisation for Research NWO under grant # 200-50-073.

Figures 1-5

References

- Haan, J., V.J. van Heuven, J.J.A. Pacilly and R. van Bezooijen (1997). An anatomy of Dutch question intonation, in J. Coerts and H. de Hoop (eds.), *Linguistics in the Netherlands 1997*, John Benjamins, Amsterdam/Philadelphia, 97-108.
- Heuven, V.J. van, Haan, J., and Pacilly, J.J.A. (1998). Global and local characteristics of Dutch questions in play-acted and spontaneous speech, *Proceedings of an ESCA Workshop on Sound patterns of spontaneous speech*, La Baume-lex-Aix, 139-142.
- Heuven, V.J. van, J. Haan, E. Janse and E.J. van der Torre (1997). Perceptual identification of sentence type and the time-distribution of prosodic interrogativity markers in Dutch, in *Proceedings of an ESCA workshop on intonation*, Athens, 317-320.
- Thorsen, N. (1980). A study of the perception of sentence intonation - Evidence from Danish, *Journal of the Acoustical Society of America* 67, 1014-1030.

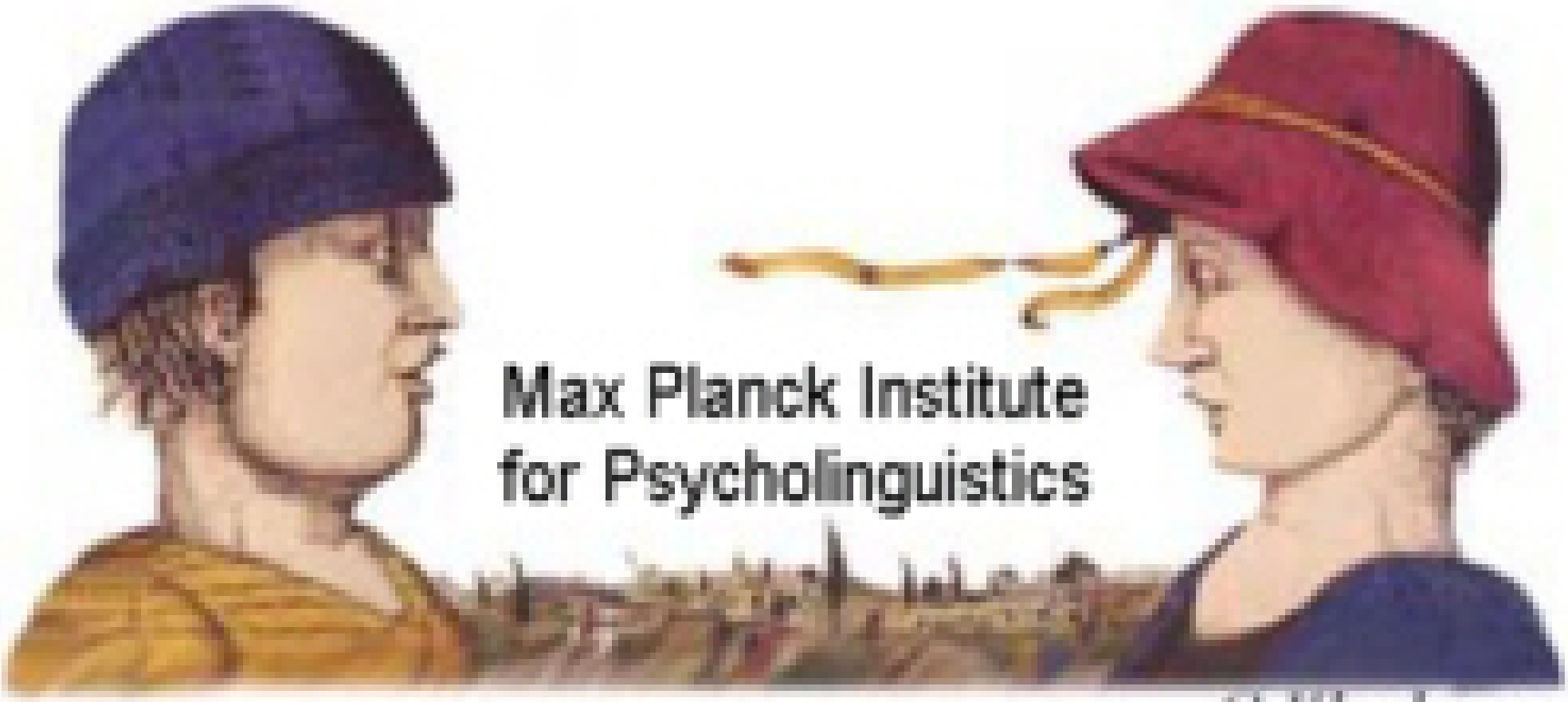

Board


Call for Papers

Sponsors

Katholieke Universiteit Nijmegen

Publication

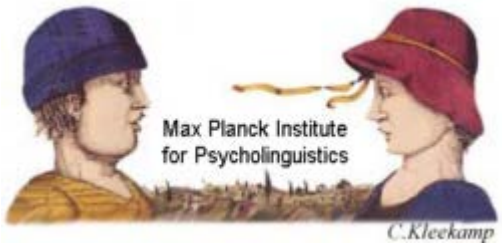

Max Planck Institute
for Psycholinguistics
C.Kleekamp

Programme

Not Found

The requested document was not found on this server.

Web Server at labphon7.ruhosting.nl



LabPhon 7

Netherlands Map

Netherlands map

Not Found

The requested document was not found on this server.

Web Server at labphon7.ruhosting.nl