It is with great pleasure that we welcome you to Cologne to PaPE 2017!

The Phonetics and Phonology in Europe (PaPE) conference series is a forum that has the aim of exploring disciplinary and interdisciplinary approaches to all areas of phonetics and phonology, with a special focus on Laboratory Phonology. This includes both theoretical and applied research, and in particular the relationship between the two. The series covers a wide variety of topics including tone and intonation, phonological theory, audiovisual prosody and gesture, language development, linguistic typology, language pathology, and language teaching. Methodologically, the conference also aims at bridging the gap between the fields of phonetics and phonology and fields such as psycholinguistics, neurolinguistics, and computational linguistics.

This is the second PaPE conference, following a highly successful first conference in Cambridge, UK, in June 2015. Prior to 2015, a series of biennial PaPI (Phonetics and Phonology in Iberia) conferences dates back to 2003. The broadening of the scope of the conference will hopefully lead to fruitful exchange in Europe and beyond.

For PaPE 2017, over 200 submissions were received, with authors from 40 countries. Of these, 95 will be presented at the conference.

Looking forward to your presentations and to stimulating discussions in and around the sessions.

Martine Grice

On behalf of the organising committee

Cologne 2017 PaPE

Organising committee:

Stefan Baumann
Anna Bruggeman
Francesco Cangemi
Martine Grice

Local organisation:

Martina Krüger
Christine Riek

We are especially grateful to our sponsors, the Deutsche Forschungsgemeinschaft (DFG), Language Science Press (LSP), the International Phonetic Association (IPA) and the Association for Laboratory Phonology (ALP).

Deutsche Forschungsgesellschaft | Association for Laboratory Phonology | Language Science Press | International Phonetic Association

SUNDAY 11 JUNE	
09:30 - 13:00	**Satellite workshop** *Statistical methods in Phonetic Sciences* Organisers Timo Roettger & Bodo Winter Neuer Senatssaal, Main Building, University of Cologne
14:30 - 18:30	**Satellite workshop** *Phonetics and Phonology in Loanword Adaptation* Organisers Silke Hamann & Klaas Seinhorst Neuer Senatssaal, Main Building, University of Cologne
MONDAY 12 JUNE	
08:30 - 09:30	Registration, KOMED Foyer
09:00 - 09:30	Opening ceremony
09:30 - 10:30	**Plenary talk 1:** Mirjam Ernestus *The roles of phonetics, phonology and the lexicon in the processing of reduced speech*
10:30 - 11:00	Coffee break
	Oral session 1: Generalisations chair Bob Ladd
11:00 - 11:30	Morgan Sonderegger, Michael McAuliffe and Hye-Young Bang *Segmental influences on F0: A large-scale study of cross-linguistic and interspeaker variability*
11:30 - 12:00	Yang Li, Adrian Leemann, Marie-José Kolly, Ricky Chan, Geraldine Kwek and Anna Jespersen *Language-dependent and language-independent perception of prominence*
12:00 - 12:30	Petra Hoedl, Jyrki Tuomainen and Victoria Knowland *On the role of attention and perceptual load in audio-visual speech perception*
12:30 - 13.15	Lunch
13:15 - 14:15	**Poster session 1** chair Anne Hermes
	Oral session 2: Variation chair Ian Maddieson
14:15 - 14:45	Rachel Smith and Tamara Rathcke *Dialectal variation in prosodic timing and prosody-segment interactions*
14:45 - 15:15	Cesko Voeten *Diachronic change and synchronic variation in Dutch vowel-/l/ sequences: The role of phonetics, phonology, and sociolinguistics*
15:15 - 15:45	Marivic Lesho *Phonetic restructuring in the vowel systems of two Cavite Chabacano dialects*
15:45 - 16:00	Coffee break
16:00 - 17:00	**Poster session 2** chair Timo Roettger
	Oral session 3: Accented speech chair Mario Vayra
17:00 - 17:30	Hannah King and Emmanuel Ferragne *The effect of ultrasound and video feedback on the production and perception of English liquids by French learners*
17:30 - 18:00	Nicholas Henriksen and Sarah Harper *L2 status affects L3 learning for the onset of acquisition: A developmental study of L1 English, L2 Spanish, and L3 Catalan*
18:00 - 18:30	Donna Erickson, Caroline Smith and Christophe Savariaux *Articulatory correlates of French and English metrical structure: Influences from L1*
TUESDAY 13 JUNE	
09:00 - 09:30	Registration, KOMED Foyer
09:30 - 10:30	**Plenary talk 2:** Bettina Braun *Pitch accent type affects stress perception: Evidence from infant and adult speech processing*
10:30 - 11:00	Coffee break

	Oral session 4: Prosodic boundaries chair David House
11:00 - 11:30	Hae-Sung Jeon and Amalia Arvaniti
	The effects of prosodic context on word segmentation in Korean
11:30 - 12:00	Shu-Chen Ou and Zhe-Chen Guo
	Is the cue of pitch rise to spoken word segmentation used in a language-specific or cross-linguistic way? A study of listeners of Taiwanese Southern Min
12:00 - 12:30	Sandrien van Ommen, Natalie Boll-Avetisyan, Saioa Larraza, Caroline Wellmann, Ranka Bijeljac-Babic, Barbara Hoehle and Thierry Nazzi
	Phonetic cues in French prosodic boundaries and infant prosodic processing
12:30 - 13.15	Lunch
13:15 - 14:15	**Poster session 3** chair Doris Mücke
	Oral session 5: Illusory vowels chair Cinzia Avesani
14:15 - 14:45	Harim Kwon and Ioana Chitoran
	Segmental and prosodic effects on perception and production of word-initial clusters
14:45 - 15:15	Sonia I. d'Apolito and Barbara Gili Fivela
	Vowel insertion in non-native consonant cluster production
15:15 - 15:45	Francisco Meneses, Sarah Ellen Johnson, Eleonora Albano and Ryan Shosted
	Is the vowel really disappearing in Brazilian Portuguese sandhi? An ultrasound study of vowel reduction
15:45 - 16:00	Coffee break
16:00 - 17:00	**Poster session 4** chair Henrik Niemann
	Oral session 6: (Co)articulation chair Ioana Chitoran
17:00 - 17:30	Özlem Ünal Logacev, Susanne Fuchs and Leonardo Lancia
	Can EPG contacts explain intraoral pressure shapes in voiced and voiceless stops in Turkish? Evidence from Generalized Additive Mixed Models
17:30 - 18:00	Fanny Guitard-Ivent and Cécile Fougeron
	Domain-initial strengthening as reduced coarticulation
18:00 - 18:30	Bob Ladd and Stephan Schmid
	Aspiration, laryngeal features, and obstruent effects on F0: Evidence from Zurich Swiss German
19:00	Conference Dinner & Party

WEDNESDAY 14 JUNE	
09:30 - 10:30	**Plenary talk 3:** Maria-Josep Solé *Language-specific adjustments to phonetic constraints and cross-linguistic patterns*
10:30 - 11:00	Coffee break
	Oral session 7: Intonation chair Jennifer Cole
11:00 - 11:30	James S. German
	Implicit social cues influence the interpretation of intonation
11:30 - 12:00	Jana Neitsch, Daniela Wochner, Katharina Zahner and Nicole Dehé
	Who likes liver? How German speakers use prosody to mark questions as rhetorical
12:00 - 12:30	Francisco Torreira and Martine Grice
	Flexibility in the association of tones captures melodic alternations in Spanish
12:30 - 13.15	Lunch
13:15 - 14:15	**Poster session 5** chair Christine Röhr
14:15 - 15:15	**Plenary talk 4:** Jonathan Barnes *Integrating pitch and time in intonational phonetics and phonology*
15:15 - 15:30	**Closing ceremony**

16:00 - 19:30 **Satellite workshop** *Introduction to the Naive Discriminative Learning package* Organiser Fabian Tomaschek
KOMED Room 313

Posters

	Session 1 Monday 12 June 13:15 - 14:15	Chair: Anne Hermes
1	Sarah Harper, Louis Goldstein and Shri Narayanan	Stylistic effects on the acoustic and articulatory properties of English rhotics
2	Andrea Peskova	Czech ToBI
3	Ishrat Rehman and Amalia Arvaniti	Perceptual similarity spaces of British English vowels by speakers of Pakistani Urdu
4	Mary Baltazani, Katerina Nicolaidis and Anastasia Chionidou	Vowel spaces in six regional Greek varieties: An acoustic and articulatory analysis
5	Eugen Klein, Jana Brunner and Phil Hoole	Flexibility of the acoustics-to-articulation mapping: Evidence from a bidirectional perturbation study
6	Mark Gibson, Stavroula Sotiropoulou, Stephen Tobin and Adamantios Gafos	Articulatory overlap in a subset of stop+lateral clusters in Spanish
7	Katharina Zahner, Heather Kember, Anne Cutler and Bettina Braun	Museum or musical? – Pitch accent type affects word recognition in Australian English
8	Thomas Kettig	Diachronically stable, lexically specific variation: The phonological representation of secondary /æ/-lengthening
9	Takeki Kamiyama and Giuseppina Turco	Non-local temporal adjustments caused by length contrasts: The case of Japanese
10	Leonardo Lancia, George Krasovitskiy and Franziska Stuntebeck	Coordinative patterns underlying speech rhythm
11	Matt Bauer	Light and dark /l/ in American English: The role of tongue narrowing
12	Beata Lukaszewicz, Ewa Zajbt, Urszula Krawczyk and Janina Molczanow	Lombard effect-related acoustic changes in the production of subsidiary stress
13	Jeffrey Lamontagne and Francisco Torreira	Production planning and frequency effects in sandhi: Hiatus resolution in Spanish
14	Olivier Crouzet and Julien Millasseau	A comparison of the information conveyed by static and dynamic formant cues for vowel classification in vowel-vowel sequences
15	Gilbert Ambrazaitis and David House	Multimodal prominences: Exploring the interplay and usage of focal pitch accents, eyebrow beats and head beats in Swedish news readings

	Session 2 Monday 12 June 16:00 - 17:00	Chair: Timo Roettger
1	Mario Ruiz Moreno and Christoph Gabriel	Voice onset time in Brazilian Portuguese-Pomerano bilinguals
2	Jalal Al-Tamimi	Acoustics of constricted epilarynx: The case of voiced coronal pharyngealized consonants in Jordanian and Moroccan Arabic
3	Elina Rubertus, Dzhuma Abakarova, Jan Ries and Aude Noiray	Comparing coarticulatory directions in child speech
4	Baris Kabak and Christina Domene Moreno	A cross-linguistic perspective on stress-meter alignment in music: Evidence from Turkish children's songs
5	Massimiliano M. Iraci, Mirko Grimaldi and Barbara Gili Fivela	Dynamic aspects of the alveolar fricative vs. stop contrast in Parkinson’s disease
6	Seung Kyung Kim	Interaction between word length and emotional prosody
7	Florence Baills, Nerea Suárez González, Santiago González-Fuente and Pilar Prieto	Gestures copying the F0 curve of lexical tones help learning Mandarin tones and words
8	Qianwen Guan	Phonological knowledge and phonetic detail in the production of stop sequences by Mandarin native speakers
9	Leonardo Barón Birchenall and Noël Nguyen	Synchronization of speech rhythm between Spanish-speaking interlocutors
10	Ingo Feldhausen and Alina Lausecker	Diatopic variation in the prosodic realization of left-dislocations in Spanish
11	Maria O'Reilly and Ailbhe Ní Chasaide	Phonatory and duration effects of focus position on the right phrase edge in South Connaught Irish
12	Bodo Winter, Márton Sóskuthy and Marcus Perlman	R is for rough: Sound symbolism for English touch concepts
13	Inyoung Kim and Maelle Amand	After Low High…Low High? A study on basic prosodic unit right boundaries in Korean
14	Aritz Irurtzun and Maia Duguine	The prosody of different question strategies in Labourdin Basque and the syntax-phonology interface

	Session 3 Tuesday 13 June 13:15 - 14:15	Chair: Doris Mücke
1	Fabian Santiago and Paolo Mairano	Do Spaniards speak faster than Mexicans? Studying Spanish rhythm in natural speech
2	Daniela Wochner and Nicole Dehé	The prosody of verb-first constructions in German: A comparison of information seeking questions, rhetorical questions and exclamatives
3	Rogelio Méndez and Corine Astésano	Perception of the downstepped final accent in French
4	Kim Strütjen, Dinah Baer-Henney, Peter Indefrey and Ruben van de Vijver	The role of perception in learning vowel nasalization
5	Sophie Egger, María Biezma and Bettina Braun	On the realization of bouletic bias in German questions
6	Bradley Rentz and Victoria Anderson	And finally, no geminates! Pohnpeian consonantal length contrasts in initial, medial, and final position
7	Shelece Easterday	The relationship between syllable structure complexity and vowel reduction processes
8	Anastasia Karlsson and David House	High boundary tones in spontaneous Southern Swedish
9	Cong Zhang	Tianjin Mandarin tunes: Production and perception data
10	Amrei Walkenhorst	Discrimination of German tense and lax vowels in German monolingual and German-Turkish and German-Russian bilingual children
11	Oana Niculescu, Ioana Vasilescu, Vieru Bianca, Lori Lamel and Martine Adda-Decker	Semi-automatic analyses of vocalic sequences in Romanian
12	Hannah Leykum and Sylvia Moosmüller	Phonotactic and morphonotactic consonant clusters in Standard German German, Standard Austrian German and Standard French
13	Renate Raffelsiefen and Anja Geumann	The phonology and phonetics of vowel trajectories in English
14	Cristel Portes and Leonardo Lancia	Acoustic and perceptual contrast between implicative and continuative intonation in French

	Session 4 Tuesday 13 June 16:00 - 17:00	Chair: Henrik Niemann
1	Meghan Dabkowski	Speech rate effects on Mexico City Spanish vowel weakening
2	Daniel McCarthy and Jalal Al-Tamimi	Place of articulation in plosives: A light on the mapping between phonetics and phonology
3	Núria Esteve-Gibert, Hélène Lœvenbruck, Marion Dohen, Thierry Legou and Mariapaola D'Imperio	The development of prosodic and gesture cues to focus in French pre-schoolers
4	Chelsea Sanker	Lexical retrieval and effects of homophones
5	Jacopo Torregrossa, Diana Dimitrova and Martine Grice	The detection of semantic incongruity on post-focal constituents in Italian
6	Michael Wagner and Michael McAuliffe	Three dimensions of sentence prosody and their (non-)interactions
7	Andreas Baumann and Kamil Kaźmierski	Lazy speakers and distant sounds: On the role of articulatory difference in phonotactic production, acquisition, and change
8	Timo B. Roettger and Mathias Stoeber	Tracking continuous motor responses as a window into real-time speech perception
9	Farhat Jabeen	Position vs. prosody: Focus realization in Urdu and Hindi
10	Valéria Krepsz and Mária Gósy	Sentence-final lengthening in Hungarian: Effects of phonemic length and word length
11	Qian Luo, Karthik Durvasula and Yen-Hwei Lin	Are consonantal effects on F0 conditioned by enhancement of contrasts of tones?
12	Margaret Zellers and Antje Schweitzer	Normalizing versus initializing cues to pitch (mis-)matching in conversation
13	Elisabeth Delais-Roussarie and Giuseppina Turco	The intonation of alternative questions in French
14	Stephen Tobin	Perceptual category adaptation: An index of cross-language coupling
15	Michelina Savino	Speaker variability and accomodation processes: The case of Bari Italian question intonation

	Session 5 Wednesday 14 June 13:15 - 14:15	Chair: Christine Röhr
1	Marianne Kusterer, Jana Neitsch, Bettina Braun and Nicole Dehé	Interpreting rhetorical questions: The influence of pitch accent type, voice quality and the modal particle *denn*
2	Jan Fliessbach	From lab speech to dialogue corpora: Givenness and Obviousness in Castilian Spanish intonation
3	Mariana Hungria and Eleonora Cavalcante Albano	The use of the upper and lower vocal tract in three Brazilian children from 0:06 to 1:07
4	Marcel Schlechtweg	Acoustic characteristics of novel English adjective-noun constructions
5	Rachel Albar and Hiyon Yoo	Perceiving and producing clusters in French: The case of Japanese learners
6	Timo Buchholz and Uli Reich	Conflicting functions and phonetic cues in bilingual Spanish and Quechua
7	Giuseppina Turco, Karim Shoul and Rachid Ridouane	On the phonetics of four-level length contrast: Evidence from Moroccan Arabic
8	Antje Stoehr, Jessica Zurlo and Janet van Hell	Perceptual voicing asymmetries in Dutch and English native listeners
9	Adèle Jatteau and Michaela Hejná	Dissimilation can be gradient: Evidence from Aberystwyth English
10	Jennifer Cole, Jose Ignacio Hualde and Suyeon Im	Imitation evidence for the encoding of prosodic detail in prenuclear accent patterns
11	Stefan Baumann, Jane Mertens and Janina Kalbertodt	The influence of information status on prenuclear accents in German
12	Natacha Chevrier	Aerodynamic and articulatory explanations for the presence of nasal consonants in oral contexts in Bribri (Chibchan, Costa Rica)
13	Kathleen Jepson, Janet Fletcher and Hywel Stoakes	Post-tonic consonant lengthening in Djambarrpuyŋu
14	Sergio Robles	Pragmatic factors and intonational overlap: Vocatives and imperatives compared
15	Yohann Meynadier, Yulia Gaydina and Antoine Giovanni	Is the voicing-dependant duration of obstruents physiological in French?

Conference Venue
KOMED im Mediapark
Im MediaPark 7, 50670 Köln
+49 221 5743 142

Contact
Stefan Baumann +49 176 23291210
Anna Bruggeman +49 157 72714161
Francesco Cangemi +49 176 31726832
Martine Grice +49 179 76464940

Poster symbol legend

Posters will have symbols on the top right hand corner to indicate subject areas.

 integration phonetics & phonology

 vowels

 consonants

 syllables

 lexical tone

 intonation

 gesture

 rhythm

 prosodic boundaries

 articulation

 perception

 lexical access

 bilingualism

 child speech

 music

 variation

 sound change

 non-lexical meaning

 social meaning

 typology

 phonological theory

 imitation

 pathology

 tools

 neurophysiology

The roles of phonetics, phonology and the lexicon in the processing of reduced speech

Miriam Ernestus
[1]*Centre for Language Studies, Radboud University*
[2]*Max Planck Institute for Psycholinguistics*

In casual conversations, many word tokens are reduced compared to their citation forms. For instance, American English *yesterday* is often pronounced like *yeshay* and French *menu* like *mnu*. Native speakers are seldom aware of these reductions and they process these effortlessly. In this talk, I address the roles of phonetics, phonology, and the lexicon in the processing of reduced speech, by comparing native with non-native listeners, among other approaches. I focus on three series of experiments.

In the first study [1], we focused on the role of phonotactic constraints. We investigated how English *can*, unreduced *can't*, and reduced *can't* with an absent or weak /t/, presented in short phrases spliced from spontaneous conversations, are understood by native listeners of languages that do (English and Dutch) or do not (Spanish and Mandarin) allow word-final /nt/ and schwa. We found direct effects of the listeners' native phonotactic constraints: low proficiency learners of English without final /nt/ in their native languages have a clear bias towards *can*, instead of *can't*. Moreover, we observed an indirect effect: whereas, for the interpretation of reduced *can't*, the Dutch and English listeners rely on the fine phonetic details cueing reduced /nt/ (residuals of /t/, quality of the preceding vowel), Spanish and Mandarin listeners hardly do so. Their low experience with word-final /nt/ and schwa in their native languages inhibits them to interpret the corresponding (subtle) phonetic cues. Native phonotactic constraints affect phonetic processing.

In a second study [2], we focused on the role of the reduced word variants' frequencies of occurrence. Previous research [e.g., 3] suggests that native listeners process a given phonological variant of a word more quickly if it is the most frequent variant. We replicated this effect of word-specific variant frequency for the absence versus presence of schwa in French words (e.g. *mnu* versus *menu*): native listeners recognize a given phonological variant (i.e., either with or without schwa) of a given word more quickly, the more frequent this variant is relative to the other variant of that word. We also tested advanced learners of French and found that their word recognition times correlate with the word variant frequencies reflecting their own language input rather than native listeners' input. This implies that the frequency effect supports the hypothesis that listeners are optimally tuned to their own language input. We argue that the frequency effects cannot be accounted for with frequency sensitive phonological reconstruction rules. Rather, the frequency effects show that reduced word variants are stored in the listener's mental lexicon.

A third study [4] also tested the role of variant frequency in word recognition. It focussed on the processing of French words ending in obstruent-liquid-schwa (e.g. *litre*). In a corpus study [5], we have shown that the obstruent may be absent due to coarticulation and articulatory weakening, rather than to a categorical (phonological or lexical) process. In [4], we show that native listeners and advanced learners of French recognize pronunciation variants without that obstruent (e.g. *li* for *litre*) more quickly than variants that occur less often and contain the obstruent (e.g. *lire* for *litre*). The role of variant frequency is thus not restricted to the processing of phonological variants but extends to phonetic variants, resulting from articulatory weakening.

These studies show the interplay of phonetics, phonology, and the mental lexicon in the processing of reduced speech. Phonetic processing is affected by phonology; the recognition of different word variants is affected by the usage based factor frequency of occurrence; and the lexicon may play an important role in the processing of phonological and phonetic word

variants. Phonetics, phonology and the mental lexicon are thus more integrated than is traditionally assumed. I will discuss how to account for our findings.

References

Ernestus, M., Kouwenhoven, H., & Mulken, M. van. 2017. The direct and indirect effects of the phonotactic constraints in the listener's native language on the comprehension of reduced and unreduced word pronunciation variants in a foreign language. *Journal of Phonetics* 62, 50-64.

Brand, S. & Ernestus, M. in press. Listeners' processing of a given reduced word pronunciation variant directly reflects their exposure to this variant: evidence from native listeners and learners of French. *Quarterly Journal of Experimental Psychology*. http://dx.doi.org/10.1080/17470218.2017.1313282

Ranbom, L. J., & Connine, C. M. 2007. Lexical representation of phonological variation in spoken word recognition. *Journal of Memory and Language*, *57*, 273–298.

Brand, S. & Ernestus, M. submitted. The use of contextual cues and of variant frequency in the comprehension of reduced speech by highly proficient language learners.

Brand, S. & Ernestus, M. under revision. Reduction of word-final obstruent-liquid-schwa clusters in casual European French.

Pitch accent type affects stress perception: evidence from infant and adult speech processing

Bettina Braun
Department of Linguistics, University of Konstanz, Germany

It is frequently claimed that stressed syllables "have higher f0, greater intensity, and longer duration than unstressed syllables" (more than 20,000 Google hits for the combination of the search terms "stressed syllable" and "higher pitch" (or "higher f0")). Yet a trochee with a rising intonation (e.g., *really?*) is an immediate counterexample, because it has a low-pitched stressed syllable followed by higher-pitched unstressed syllable.

Intonation languages (in this talk I focus on German, Dutch and English) signal paralinguistic and pragmatic information by means of different types of pitch accents and boundary tones, which results in different alignments of high and low tonal targets with regard to lexically stressed syllables. Due to this variability in phrase-level intonation, f0 (and in particular the position of high tonal targets) cannot be considered a reliable cue to the position of lexical stress in these languages. So the above claim of higher f0 in stressed syllables is only correct in very specific contexts, e.g., when target words act as focus in statements with a final fall (or are produced in isolation). Results from production studies with linguistically diverse speech materials show that f0 is not an acoustic correlate of stress (e.g., Sluijter & Van Heuven 1996b, 1996a, Dogil 1995, Szalontai, Wagner, Mády & Windmann 2016, Kochanski, Grabe, Coleman & Rosner 2005). In perception, however, German and English listeners rely on f0 to identify lexical prominence or stress (Kohler 2008, Fry 1958). Is this conclusion also an artefact of lacking linguistic (or maybe methodological) diversity?

In this talk, I will first briefly summarize recent perception data from artificial language studies (Bion, Benavides-Varela & Nespor 2011, Abboub, Boll-Avetisyan, Bhatara, Höhle & Nazzi 2016). Next, I will present experimental evidence from offline (stress identification) and online experiments (headturn-preference paradigm, eye-tracking) from our labs. Results show that German 9-month olds treat high pitch as a necessary cue for lexical stress and metrical segmentation: they do not extract trochees with a low-pitched stressed syllable but only with a high-pitched stressed syllable (Zahner, Schönhuber & Braun 2016). For German adults, low-pitched stressed syllables lead to more errors in an offline stress identification task (Egger 2015) and to stronger activation of stress competitors in visual-world eye-tracking studies compared to high-pitched stressed syllables (Zahner, Schönhuber, Grijzenhout & Braun 2016b). The latter finding was also replicated for English (Zahner, Kember & Braun submitted). Neurophysiological data for German point in the same direction (Friedrich, Alter & Kotz 2001). In sum, high pitch is temporarily interpreted as stress in all the experimental paradigms. Pitch immediately affects online speech processing.

The talk finally discusses potential mechanisms that can account for the observed f0-stress interference effects in perception, such as the salience of the f0 cue (e.g., Wagner, Cwiek & Samlowski 2016, Baumann 2014) and the frequent occurrence of high-pitched stressed syllables in both infant-directed and adult-directed speech (Dainora 2006, Peters, Kohler & Wesener 2005, Zahner, Schönhuber, Grijzenhout & Braun 2016a). In closing, I will briefly touch upon implications for lexical acquisition and phrase-level prominence judgments.

References

Abboub, Nawal, Natalie Boll-Avetisyan, Anjali Bhatara, Barbara Höhle & Thierry Nazzi. 2016. An exploration of rhythmic grouping of sequences by French- and German-learning infants. *Frontiers in Human Neuroscience 10*. 292.

Baumann, Stefan. 2014. *The importance of tonal cues for untrained listeners in judging prominence*. Paper presented at the ISSP 2014, Cologne, Germany.

Bion, Ricardo A. H., Silvia Benavides-Varela & Marina Nespor. 2011. Acoustic markers of prominence influence infants' and adults' segmentation of speech sequences. *Language and Speech 54*(1). 123-140.

Dainora, Audra. 2006. Modeling intonation in English: A probabilistic approach to phonological competence. In L. Goldstein, D. Whalen & C. Best (Eds.), *Laboratory Phonology 8* (pp. 107 – 132). Berlin, New York: Mouton de Gruyter.

Dogil, Grzegorz. 1995. Phonetic correlates of word stress. *Arbeitspapiere des Instituts für Maschinelle Sprachverarbeitung 2*. 1-60.

Egger, Sophie. 2015. *The impact of pitch accents on the identification of word stress in German*. Unpublished MA Thesis, University of Konstanz.

Friedrich, Claudia, Kai Alter & Sonja A. Kotz. 2001. An electrophysiological response to different pitch contours in words. *Neuroreport 12*(15). 3189-3191.

Fry, D. B. 1958. Experiments in the perception of stress. *Language and Speech 1*. 126-152.

Kochanski, Gregory P., Esther Grabe, John Coleman & Burton S. Rosner. 2005. Loudness predicts prominence: fundamental frequency lends little. *Journal of the Acoustical Society of America 118*(2). 1038-1054.

Kohler, Klaus J. 2008. The perception of prominence patterns. *Phonetica 65*. 257-269.

Peters, Benno, Klaus Kohler & T. Wesener. 2005. Melodische Satzakzentmuster in prosodischen Phrasen deutscher Spontansprache - Statistische Verteilung und sprachliche Funktion [Melodic sentence accent patterins in spontaneous German prosodic phrases - statitical distribution and linguistic function]. In Klaus Kohler, Felicitas Kleber & Benno Peters (Eds.), *Prosodic Structures in German Spontaneous Speech (AIPUK 35a)* (pp. 185-201). Kiel: IPDS.

Sluijter, Agaath M.C. & Vincent J. Van Heuven. 1996a. *Acoustic correlates of linguistic stress and accent in Dutch and American English*. Paper presented at the 4th International Conference on Spoken Language Processing (ICSLP), Philadelphia.

Sluijter, Agaath M.C. & Vincent J. Van Heuven. 1996b. Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America 100*(4). 2471-2485.

Szalontai, Ádám, Petra Wagner, Katalin Mády & Andreas Windmann. 2016. *Teasing apart lexical stress and sentence accent in Hungarian and German*. Paper presented at the Phonetik und Phonologie im deutschsprachigen Raum, Munich, Germany.

Wagner, Petra, A. Cwiek & Barbara Samlowski. 2016. *Beat it! Gesture-based prominence annotation as a window to individual prosody processing strategies*. Paper presented at the Phonetik und Phonologie im deutschsprachigen Raum, Munich, Germany.

Zahner, Katharina, Heather Kember & Bettina Braun. submitted. Mind the peak: When museum is temporarily understood as musical in Australian English

Zahner, Katharina, Muna Schönhuber & Bettina Braun. 2016. The limits of metrical segmentation: Intonation modulates the extraction of embedded trochees. *Journal of Child Language 43*(6). 1338-1364.

Zahner, Katharina, Muna Schönhuber, Janet Grijzenhout & Bettina Braun. 2016a. *Konstanz prosodically annotated infant-directed speech corpus (KIDS Corpus)*. Paper presented at the 8th International Conference on Speech Prosody, Boston, MA.

Zahner, Katharina, Muna Schönhuber, Janet Grijzenhout & Bettina Braun. 2016b. *Pitch accent type affects lexical activation in German: Evidence from eye tracking*. Paper presented at the 16th Speech Science and Technology Conference (SST 2016), Sydney, Australia.

Language-specific adjustments to phonetic constraints and cross-linguistic patterns

Maria-Josep Solé
Universitat Autònoma de Barcelona

It is generally accepted that cross-linguistically common phonological patterns partly derive from physiological and physical phonetic constraints which fail to be normalized by the listener (Ohala 1981) or shift the modal values of the phonetic category (Bybee 2012, Garrett and Johnson 2013). We propose that if languages do not resist these constraints, they tend to show up as 'phonetically natural' sound patterns or universal tendencies. But languages can moderate the constraints of the speech production system in a variety of ways. They may circumvent them or oppose them by taking active countermeasures and other less common (but also 'phonetically natural') patterns may result. Thus, the claim made in this paper is that phonetic constraints are not inevitable but may be moderated by language-specific adjustments.

In this talk we investigate further how the difference between a common sound pattern and a less common pattern may stem from obeying or resisting (in a variety of ways) physical phonetic constraints. We will first review experimental and modelling data illustrating different ways in which languages may deal with the 'Aerodynamic voicing constraint' (AVC, Ohala 1981, 2011) which impacts negatively on obstruent voicing. Languages may yield to the constraint and passively devoice stops leading, for example, to final obstruent devoicing or utterance-initial neutralization of the stop voicing contrast. Alternatively, they may resist the constraint and avoid devoicing. If voicing during the stop closure is to be maintained, languages have available a limited set of articulatory adjustments operating along different parameters (nasal or oral resistance; closure duration; larynx lowering; tongue body lowering, etc.). Such limited set of parameters result in broad linguistic patterns that recur frequently, e.g., *emerging nasals* (see example 1), *spirantization* (2), *d-flapping/lateralization* or *retroflexion* (3), and *implosivization* (4) (Sprouse, Solé & Ohala 2008, Ohala 2011, and Solé 2012, 2014). Indeed within each of these patterns languages may vary in the quantitative values used along each dimension, resulting in fine phonetic differences, for example, in amount of voicing, degree of spirantization, etc.

A second example is changes in the timing of the nasal and oral gestures; such differences in the phasing of the gestures, in turn, may facilitate or prevent the transglottal flow for voicing. We analyze data showing that postnasal *voicing* (i.e., the voicing of obstruents after a nasal (example 5) is a commonly found and phonetically grounded sound pattern. However, languages may counteract the prolonged transglottal flow and voicing into the stop closure in NC clusters by an early raising of the velum relative to the oral closure. This is found in Shekgalagari and Tswana (Solé, Hyman & Monaka 2010) which, instead, show postnasal *devoicing* (example 6). (Early velum raising in these languages also accounts for postnasal affrication). In a similar vein, late velic closure (due to the sluggishness of the velum) is at the origin of *nasal assimilation* (example 7), whereas early velic closure is at the root of *pre-stopped nasals* and *denasalization* (examples 8 and 9 respectively).

The reviewed data suggest that phonetic constraints are not unavoidable and that common phonological patterns may stem from ways in which languages deal with physical phonetic constraints.

Examples cited:

(1) Emergence of non-etymological nasals.

Nasalization of voiced stops, dialects of Tai (Li 1977:68-69; 107)

Proto-Tai	*Siamese*	*Lungchow*	*Po-ai*	
**ʔb-*	baa	baa	**m**aa	'shoulder'
**ʔd-*	dïat	dïït	**n**aat	'hot'

(2) Spirantization

Basaa (Teil-Dautrey 1991)

PB *gubú ŋ-gù**β**í 9/10 'hippopotamus'; PB *tòpè n-dɔ**β**ɔ 9/10 'mud'

(3) d-flapping/lateralization/gliding (a) or retroflexion (b)

a. Palenquero /juˈda/ [juˈ**l**a] ~ [ju**ð̞**ˈa] ~ [juˈ**ɾ**a] 'to help' (Piñeros 2003: 1191)

b. Apical voiceless stops are dental /t̪ ⁿt̪/ and voiced stops are retroflex [ɖ̪ ⁿɖ̪] in the Proto-Oceanic consonant system (Ozanne-Rivierre 1992).

(4) Implosivization

Distribution of plain voiced stops and implosives in Konde (Greenberg 1970)

/b d/ → [b d]/ N__

[ɓ ɗ] elsewhere e.g., ulu-ɓafu 'rib' - imbafu 'ribs'

(5) Postnasal voicing

Southern Italian *campo* [ˈkam**b**o] 'country', *dente* [ˈdɛn**d**e] 'tooth', *bianco* [ˈbjaŋ**g**o] 'white' (Rohlfs 1966: 363)

(6) Postnasal *devoicing*

Shekgalagari [xʊ balɛla] 'to count for' , [m̩ paléla] 'count for me!' (Solé et al. 2010)

(7) Nasal assimilation

/nt/, /nd/ > /nn/ in Ganda, Mabuumbi and Fante

(8) Pre-stopped nasals

-VN > -VDN languages of South East Asia (Blust 1991:149)

(9) Denasalization

/nn/>/nd/ in Kikongo and Kiyaka,

[na] > [n^da] >[^{n}da] > [da] in some Athabaskan languages (Leer 1996)

References

Bybee, J. 2012. Patterns of lexical diffusion and articulatory motivation for sound change. In M.J. Solé & D. Recasens (eds.), *The Initiation of Sound Change* (pp. 211-234). Amsterdam: John Benjamins.

Garrett, A. and Johnson, K. 2013. Phonetic bias in sound change. In Alan C. L. Yu (ed.), *Origins of sound change: Approaches to phonologization* (pp. 51-97). Oxford: OUP.

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 the Parasession on Language and Behavior* (pp. 178 -203). Chicago: Chicago Linguistic Society.

Ohala, J.J. 2011. Accommodation to the aerodynamic voicing constraint and its phonological relevance. *Proceedings of the 17th ICPhS*, Hong Kong, 64-67.

Solé, M.J. 2012. Natural and Unnatural Patterns of Sound Change? In M.J. Solé & D. Recasens (eds.) *The Initiation of Sound Change* (pp. 123-146). Amsterdam: John Benjamins.

Solé, M.J. 2014. The perception of voice-initiating gestures. *Laboratory Phonology* 5.1, 37-68.

Solé, M.J., Larry M. Hyman & Kemmonye C. Monaka. 2010. More on post-nasal devoicing: The case of Shekgalagari. *Journal of Phonetics* 38.4, 604-616.

Sprouse, R., Solé, M.J., & Ohala, J.J. 2008. Oral cavity enlargement in retroflex stop. *Proceedings of the International Seminar on Speech Production 2008*. Strasbourg, France, 425- 428.

Integrating pitch and time in intonational phonetics and phonology

Jonathan Barnes
Boston University

The relationship between the phonological encoding of contrasting sound categories in human language and the constellations of phonetic properties realizing them in the speech signal is an intensively researched, and also hotly disputed area of inquiry. At issue is the traditional notion of a more-or-less direct, one-to-one mapping between distinctive feature specifications and their stable, universal phonetic expressions. Some have argued, for example, that this account is too simple for what increasingly looks more like complex webs of integrating cues varying greatly in presence and reliability from language to language, or even speaker to speaker, but whose general character, in terms of the properties that integrate, may be stable enough to be considered definitional (Kingston et al. 2008).

But if questions of cue integration and feature definition are still contested in segmental phonology, they are all but unbroached in the area of prosody and intonation. Researchers increasingly appreciate the range of factors interacting in the realization of tonal contrasts (e.g., intensity, voice quality, segmental spectra, F0 contour shape, etc.). Still, autosegmental representations predispose us to focus, often disproportionately, on the two properties of intonation contours that the model does encode directly: F0 scaling and timing. These latter are the focus of this talk as well. Although more-or-less orthogonal dimensions in AM models, scaling and timing are known to interact in complex and still poorly understood ways. One recent example is a demonstration that the Kappa effect (Cohen et al. 1953) can influence English speakers' perception of prosodic boundaries (Brugos 2015). The Kappa effect is a perceptual illusion whereby distance travelled in some domain influences estimation of time elapsed along the journey. It has, furthermore, a sibling in the perception literature known as the Tau effect, essentially the mirror image: differences in elapsed time cause over- or underestimations of distance travelled during a given interval (Helson 1930). In prosody this calls to mind, for example, cases in which the timing of F0 peaks in various languages interacts, in either perception or production, with peak height, such that later peaks are equated in various ways with higher ones. Gussenhoven's 2004 explanation of this pattern in the context of his Biological Codes is conceptually quite similar to the Tau effect. The opposite pattern, however, whereby high is suggested not by later peaks, but by early, has also been documented (Gussenhoven et al. 2016, Cangemi et al. 2016).

Here we present the results of two experiments investigating these interactions in the context of peak height and timing covariation documented in Gothenburg Swedish (Segerup & Nolan 2006). We show that the Gothenburg pattern of perceptual integration of peak timing and height obtains equally well for American English speakers, suggesting that process is therefore not experience-driven, but rather the result of some deeper aspect of auditory processing (Lee & Katz 2016). While here we might invoke the Tau effect, other possibilities exist as well. Our second experiment argues in favor of one such alternate account: Specifically, an approach based on the idea that speakers integrate F0 information over time, in a manner modeled as a weighted average, which we have called in other contexts the Tonal Center of Gravity (Barnes et al. 2012). TCoG is conceived as a perceptual reference location for F0 events in two dimensions: scaling on the one hand, and timing on the other. Integrating pitch and temporal information thusly yields a representation for listeners both of when, and at what pitch level, an F0 event occurred, and is at least consonant with the phonological representations of AM theory, linking tonal autosegments to particular host locations in the segmental skeleton. We suggest that this approach may provide an account of putative counterexamples to the "late equals high" equation in other languages. It may also provide the beginning of what is sure to be a far more complex picture of how the multitudinous and

varying phonetic properties of utterances integrate to yield stably contrasting categories in intonational phonology.

References

Alejna Mari Brugos. (2015). The interaction of pitch and timing in the perception of prosodic grouping. Dissertation. Boston University.

Cangemi, F., D. El Zarka, S. Wehrle, S. Baumann & M. Grice (2016). Speaker-specific intonational marking of narrow focus in Egyptian Arabic. Proceedings of the 8th Speech Prosody Conference, May-June 2016, Boston. 1-5.

Cohen, J., Hansel, C. E. M., & Sylvester, J. D. (1953, November 14). A new phenomenon in time judgement. Nature, 172, 901.

Gussenhoven, Carlos (2004) The Phonology of Tone and Intonation. Cambridge: Cambridge University.

Gussenhoven, Carlos, Lu Wang & Louis Hendrix. (2016). Yuhuan Wu tone sandhi and tone contrast maintenance. Speech Prosody 2016, 1052-1055.

Helson, H. (May 23, 1930). "The tau effect -- an example of psychological relativity". Science. 71 (1847): 536–7

Jonathan Barnes, Nanette Veillieux, Alejna Brugos & Stefanie Shattuck-Hufnagel. (2012). Tonal Center of Gravity: A global approach to tonal implementation in a level-based intonational phonology. In: Laboratory Phonology 3(2)

Kingston, J., Diehl, R. L., Kirk, C. J., and Castleman, W. A. (2008). On the internal perceptual structure of distinctive features: The [voice] contrast, Journal of Phonetics, 36, 28-54.

Lee, Sarah and Jonah Katz. 2016. Perceptual integration of acoustic cues to laryngeal contrasts in Korean fricatives. Journal of the Acoustical Society of America 139, 605

Segerup, M. & Nolan, F. 2006. Gothenburg Swedish word accents: a case of cue trading? In: Bruce, G., Horne, M. (eds), Nordic Prosody: Proceedings of the IXth Conference. Frankfurt am Main: Peter Lang, pp. 225-233.

Segmental influences on F0: a large-scale study of cross-linguistic and interspeaker variability

Morgan Sonderegger[1], Michael McAuliffe[1], Hye-Young Bang[1]
[1]*McGill University*

Introduction: One place the relationship between phonetics and phonology is particularly striking is sound change, which commonly involves phonologization of *phonetic precursors*: small phonetic effects assumed to hold across languages and individuals (e.g. Hombert et al. 1979). Yet relatively little is known about the *robustness* of most phonetic precursors: variability in their effect size across languages and speakers, which matters for establishing which precursors are robust enough to plausibly lead to change. Establishing the robustness of phonetic precursors is difficult because it requires large samples of comparable data from many speakers and languages. Two widely-studied precursors, which also form a good test case for an automated analysis, are the effect of vowel height and preceding consonant [voice] on F0 (*VF0* and *CF0*; Kingston, 2007). We assess the degree of cross-linguistic and interspeaker variability in VF0 & CF0 effects across 14 languages, using large corpora of read speech.

Data/Analysis: We examined read sentences in 13 languages from *GlobalPhone*—Croatian, French, German, Hausa, Korean, Mandarin, Polish, Russian, Spanish, Swedish Thai, Turkish, Vietnamese ($\sim$20 hrs each)—and English (using *LibriSpeech*; $\sim$2 hrs), all force-aligned using the Montreal Forced Aligner. We then extracted F0 contours in semitones (using Praat) for all {a, i, u} vowel tokens in each language in utterance-initial obstruent-vowel syllables, and excluded unreliable F0 measures. The resulting datasets contain 0.3k vowels (English)/2.4–7k vowels (other languages) from 67–113 speakers/language. Two linear mixed-effects models were fit for each language of mean vowel F0 over (1) the whole vowel and (2) its first 50 ms, to assess VF0 and CF0. Each model contained fixed effects including laryngeal class, vowel height, and various controls; and by-speaker/word random effects including by-speaker random slopes for height/laryngeal class (for VF0/CF0 models), to assess interspeaker variability.

Results: Fig. 1 shows the estimated VF0 effect size for each language and CF0 effect size for the 12 languages with 2-way laryngeal contrasts. The VF0 effect is positive in each language (high>low V), confirming the near-universality of VF0 effects (e.g. Whalen & Levitt, 1995), but its *magnitude* varies greatly across languages and does not always reach significance (4/14 languages). The smallest effects are for languages that use F0 contrastively (Thai, Vie, Man, Swe, Cro) in line with Connell (2002). The CF0 effect is positive in each language ([-voice]>[+voice] obstruent), regardless of how [voice] is phonetically implemented (c.f. Kingston & Diehl, 1994). CF0's magnitude varies greatly across languages, but the effect almost always reaches significance. Fig. 2 shows the estimated inter-speaker spread of VF0 and CF0 effects. Within-language interspeaker variability is generally large for both CF0 and VF0, enough for some speakers to show small or reversed effects. Also, the range of speakers values overlaps 0 more often for VF0 than for CF0 (12/14 vs. 7/12 languages), and languages often show more VF0 than CF0 variability.

Discussion The existence of VF0 and CF0 effects are relatively robust across languages, confirming that they are possible phonetic precursors to sound changes. Their robustness across speakers is less clear: VF0 and CF0 effects show large interspeaker variability, and (tentatively) more for VF0. This may help explain why *neither* VF0 or CF0 effects from the preceding C often phonologize (Kingston, 2011), and possibly why VF0 in particular almost never does (Hombert, 1977). A methodological finding of this study is that VF0 and CF0 effects can be detected in non-laboratory speech with minimal statistical controls, despite not accounting for many factors greatly affecting F0 (e.g. intonation, stress), suggesting that fully-automated large-scale studies of phonetic precursors involving F0 are tenable.

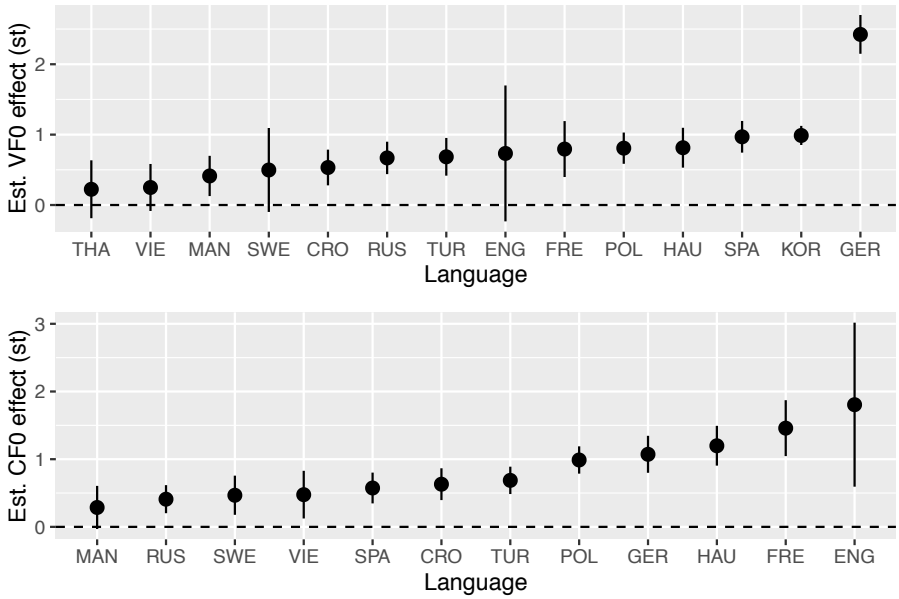


Figure 1. *Model-predicted VF0 effect (high vowel/low vowel difference; top) and CF0 effect ([-voice] obstruent/[+voice] obstruent difference; bottom) for each language, in semitones. Errorbars=95% confidence intervals ($\hat{\beta}\pm$1.96 SE). Non-overlap of the errorbars with 0 corresponds to statistical significance ($|t| > 2$).*

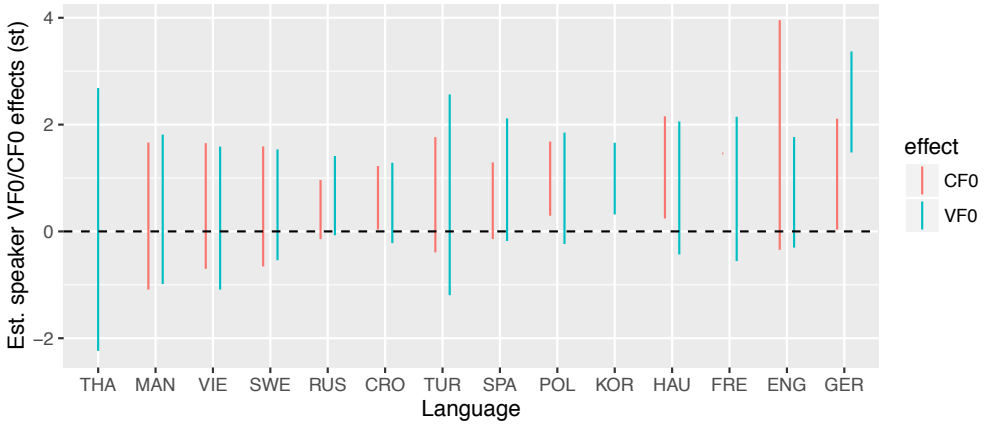


Figure 2. *Model-predicted VF0 effects and CF0 effects , across speakers of each language, in semitones. Vertical bars = predicted range of values for 95% of speakers ($\hat{\beta} \pm 1.96 \times \sigma_{\textit{by-speaker slope}}$).*

[1] Hombert, J-M., Ohala, J., & Ewan, W. 1979. Phonetic explanations for the development of tones. *Language* 55(1), 37–58.

[1] Kingston, J. 2007. F0 Segmental influences on F0: automatic or controlled? In Gussenhoven, C., & Riad, T. (Eds.), *Tones and Tunes vol. 2*. Berlin: Mouton de Gruyter, 171–210.

[2] Whalen, D., & Levitt, A. 1995. The universality of intrinsic F0 of vowels. *Journal of Phonetics* 23(3), 349–366.

[3] Connell, B. 2002. Tone languages and the universality of intrinsic F0: evidence from Africa. *Journal of Phonetics* 30(1), 101–129.

[4] Kingston, J., & Diehl, R. 1994. Phonetic Knowledge. *Language* 70(3), 419–454.

[5] Kingston, J. 2011. Tonogenesis. In van Oostendorp, M., Ewen, C., Hume, E., & Rice, K. (Eds.), *The Blackwell companion to phonology*. Oxford: Wiley-Blackwell, 2304–2333.

[6] Hombert, J-M. 1977. Development of tones from vowel height? *Journal of Phonetics* 5, 9–16.

Language-dependent and language-independent perception of prominence

Yang Li[1], Adrian Leemann[2], Marie-José Kolly[3], Ricky Chan[2],
Geraldine Kwek[1], Anna Jespersen[4]
[1]*University of Cambridge,* [2]*Lancaster University,*
[3]*University of Zurich,* [4]*Aarhus University*

Listeners of different languages vary in their perception of syllable prominence. Early research in this area has shown, for instance, that Estonian listeners were more responsive to duration cues than English listeners [1]. This may be attributable to the greater role duration plays in the Estonian quantity system, tying prosodic structures of listeners' variety to prominence perception. On the other hand, findings abound on the potentially language-independent tendencies [2] to link acoustic properties, such as higher pitch and longer duration, to perceived prominence [3], [4]. The interplay of these trends calls for a study, using unified procedures and stimuli, of the acoustics cues in prominence perception for speakers of typologically distinct and related languages. In providing this, we also test whether naïve listeners can perceive syllable prominence given limited acoustic cues and relatively vague instructions [5], [6], especially in languages where phonological prominence carries little relevance, e.g. Cantonese [7] (cf. [8]).

The experiment tests 8 varieties: French, Cantonese, Fuzhou Min Chinese, Singapore English, Southern British English, Danish, and Valais and Schaffhausen (Eastern) Swiss German, with 170 listeners in total. The base stimuli were a disyllabic logatome <papa> from a male voice, with identical durations, and pitch and intensity contours on both syllables [9], [10]. The duration, amplitude envelope and pitch contours of both syllables are then systematically manipulated through re-synthesis, e.g. in the duration group, three stimuli (L1, L2, L3) undergo stepwise increase in duration on the first syllable and decrease on the second, while three other stimuli undergo the reverse (R1, R2, R3). These then form with the base stimuli the 7 experimental levels. The same procedure is used to create 7 levels of f0 and intensity manipulations. During experiments, the stimuli were played to the listeners in random order within each block (f0, intensity, duration), and listeners were asked to identify which syllable they heard as "stronger" and to make a snap decision.

Altogether, 17850 response were collected (3 manipulations * 7 levels * 5 repetitions * 170 listeners) and plotted in Figures 1-3. In lieu of detailed findings given space constraints, we sketch the following generalisations: 1) Languages which have predictable stress (Fuzhou, French), lexical stress (all Germanic varieties tested), and no obvious stress (Cantonese) can all use acoustic cues to detect prominence at a fairly low level of processing As opposed to e.g. , [11]; 2) There are clear cross-linguistic differences in sensitivity to changes in acoustic cues. For example, Danish listeners seem to respond readily to f0 and intensity manipulations, while oblivious to changes in duration, an effect we attribute to the potential prominence of *stød* syllables, which are shorter than normal [12]. 3) Listeners of most languages exhibit a bias towards the first syllable as the site of prominence, understandable given tendencies towards iambic stress in the Germanic varieties [13] and listeners' compensation for expected final lengthening, while the reverse is seen in strongly iambic Fuzhou Chinese [14].

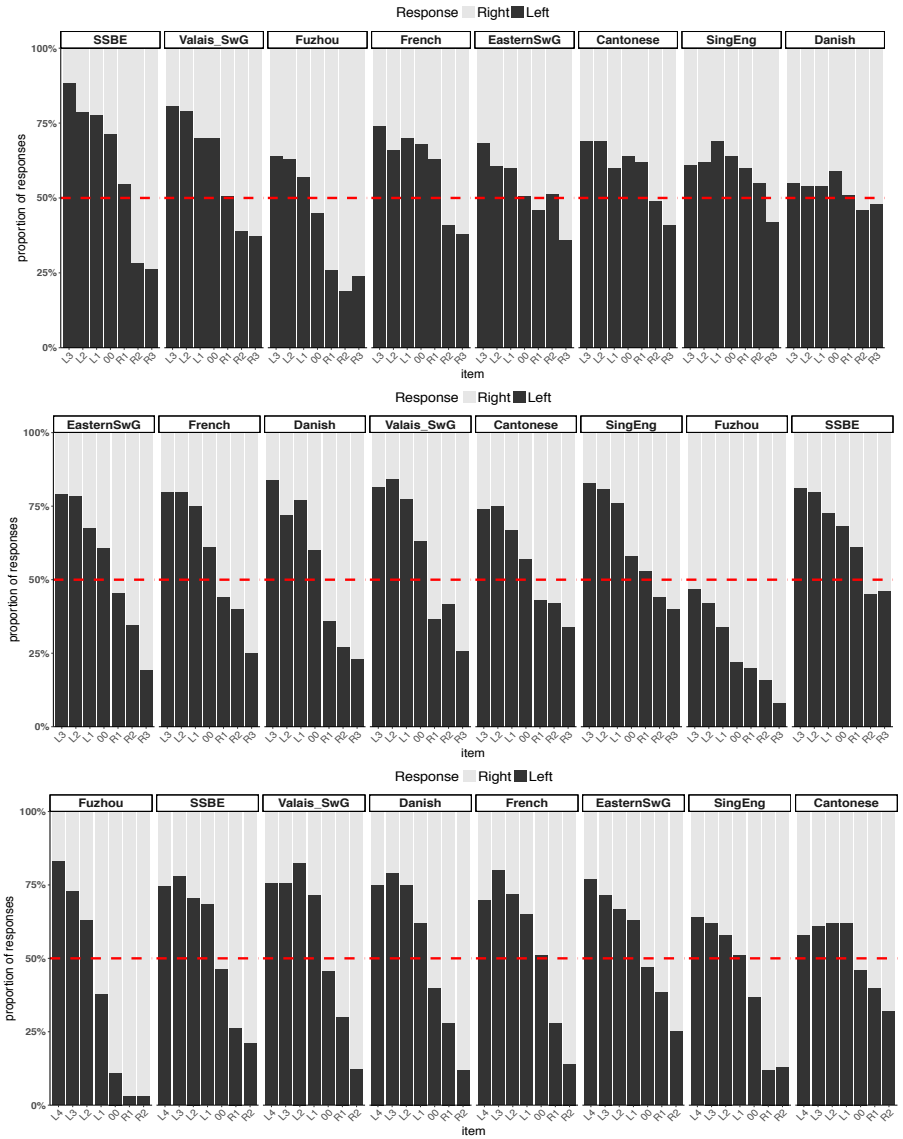


Figure 1-1. *Percentages of responses of left/right syllable as "stronger" as a function of* ***duration, intensity, and F0*** *changes (in that order). From L1 to R3, left-to-right duration/intensity/f0 ratios gradually rise. The red dotted line marks 50%.*

References

[1] I. Lehiste and R. A. Fox, "Perception of prominence by Estonian and English listeners," *Language and Speech*, vol. 35, no. 4, pp. 419–434, 1992.

[2] J. Vaissière, "Language-independent prosodic features," in *Prosody: Models and measurements*, Springer, 1983, pp. 53–66.

[3] A. Rietveld and C. Gussenhoven, "On the relation between pitch excursion size and prominence," *Journal of Phonetics*, vol. 13, pp. 299–308, 1985.

[4] B. M. Streefkerk, *Prominence: Acoustic and lexical/syntactic correlates*. LOT, 2002.

[5] C. Gussenhoven, "Does phonological prominence exist?," *Lingue e linguaggio*, pp. 7–24, 2015.

[6] C. Odé, "On the perception of prominence in Indonesian," in *Experimental Studies of Indonesian Prosody*, C. Odé and V. J. Van Heuven, Eds. Leiden: RUL, 1994.

[7] W. Y. P. Wong, M. K. Chan, and M. E. Beckman, "An autosegmental-metrical analysis and prosodic annotation conventions for Cantonese," in *Prosodic typology: The phonology of intonation and phrasing*, S.-A. Jun, Ed. Oxford: Oxford University Press, 2006, pp. 271–300.

[8] M. Yip, "Prosodic Morphology in Four Chinese Dialects," *Journal of East Asian Linguistics*, vol. 1, no. 1, pp. 1–35, Jan. 1992.

[9] K. J. Kohler, "The perception of prominence patterns," *Phonetica*, vol. 65, no. 4, pp. 257–269, 2008.

[10] A. Leemann, M.-J. Kolly, Y. Li, R. Chan, G. Kwek, and A. Jespersen, "Towards a typology of prominence perception: the role of duration," in *Proceedings of Speech Prosody 2016*, 2016.

[11] E. Dupoux, N. Sebastián-Gallés, E. Navarrete, and S. Peperkamp, "Persistent stress 'deafness': The case of French learners of Spanish," *Cognition*, vol. 106, no. 2, pp. 682–706, Feb. 2008.

[12] N. Grønnum, "F0, voice quality, and danish stød revisited," in *ICPhS 2015*, 2015.

[13] M. Trommelen, W. Zonneveld, and M. Jessen, "Word stress in West Germanic languages," in *Word prosodic systems in the languages of Europe*, H. van der Hulst, Ed. Berlin; New York: Mouton de Gruyter, 1999, pp. 478–515.

[14] Y. Li, "Tone sandhi and tonal coarticulation in Fuzhou Min," in *ICPhS 2015*, 2015, paper 0939.

On the role of attention and perceptual load in audio-visual speech perception

Petra Hödl[1], Jyrki Tuomainen[2] and Victoria Knowland[3]
[1]University of Graz, [2]University College London, [3]University of York

During a conversation we usually not only hear our interlocutors but we also see them. Hence, we receive auditory as well as visual information. This visual information is important to gain basic knowledge about the overall characteristics of the communicative situation, but beyond this visual cues are crucial for the processing of the speech sounds themselves. Seeing the mouth movements of a talker helps us to extract important phonetic cues, especially under adverse listening conditions (see e.g. [1], [2]). However, it seems that visual speech information does not simply "assist" auditory speech perception. Speech perception itself is rather commonly seen as being multisensory to its core. Support for this claim comes – amongst others – from the well-known *McGurk-illusion* (see [3]). Listeners tend to process lip-reading cues even if the auditory input itself is perfectly intelligible and sufficient to form a clear percept (see e.g. [4]). Hence, it seems somewhat "unavoidable" to process visual information during speech processing. The obvious question that arises from such findings is whether the extraction of visual speech cues is indeed an obligatory and automatic process or whether it can also be affected by top-down processes, such as attention.

Indeed, there is evidence that attention modulates the integration of acoustic and visual information during speech perception (see e.g. [5], [6], [7]). However, the exact mechanisms of this integration are not yet completely understood.

This study aimed at examining the effects of *attention* together with *perceptual load* on audio-visual speech perception. For this purpose, eighteen native speakers of British English were presented with bimodal and unimodal videos of a native talker who produced either the syllables [ba], [da] or [ga]. Around the nose of the talker a circle of geometrical shapes was displayed. Participants simultaneously performed a primary (non-linguistic) visual search task, for which they had to spot a certain geometrical shape, and a syllable identification task. The perceptual load of the primary task was manipulated and participants were exposed to three different load conditions: *no visual search task* (i.e. single task condition) for the baseline, *low load* and *high load* (i.e. dual task conditions). At the same time participants also had to report which syllables they have heard.

Following the assumptions of the *perceptual load theory* (e.g. [8], [9], [10]) we hypothesised that the processing of speech-related mouth movements should be hampered under high perceptual load. According to this theory, perception is thought to be a limited process which proceeds "automatically" – but only until it runs out of capacity. For our case we predicted that high perceptual load should yield to the perception of audio-visual stimuli more similar to the perception of auditory-only stimuli.

Our results showed an overall effect of attention. The ability to extract visual information during speech perception was reduced during the dual task condition (visual search task plus syllable identification task) compared to the single task condition (syllable identification task only). However, no convincing additional effect of perceptual load (low load vs. high load) was found.

We concluded that speech-related mouth movements might probably have a special status in perception. Their socially meaningful nature might lead to a certain amount of resistance against manipulations of perceptual load. That being said, there was also evidence for quite substantial individual differences across listeners regarding their processing of visual cues and their susceptibility to manipulations of perceptual load during speech perception.

[1] Green, K.P. & Kuhl, P.K. 1989. The role of visual information in the processing of place and manner features in speech perception. *Perception & Psychophysics*, 45, 34–42.
[2] MacLeod, A. & Summerfield, Q. 1987. Quantifying the contribution of vision to speech perception in noise. *British Journal of Audiology*, 21, 131–141.
[3] McGurk, H. & MacDonald, J. 1976. Hearing lips and seeing voices. *Nature*, 264, 746–748.
[4] Reisberg D., McLean J. & Goldfield, A. 1987). Easy to hear but hard to understand: A lipreading advantage with intact auditory stimuli. In Dodd, B. & Campbell, R. (Eds.), *Hearing by eye: The psychology of lip-reading*. London: Erlbaum, 97–114.
[5] Navarra, J., Alsius, A., Soto-Faraco, S. & Spence, Ch. 2010. Assessing the role of attention in the audiovisual integration of speech. *Information Fusion*, 11, 4–11.
[6] Alsius, A., Navarra, J., Campbell, R. & Soto-Faraco, S. 2005. Audiovisual integration of speech falters under high attention demands. *Current Biology*, 15, 839–843.
[7] Andersen, T.S., Tiippana, K., Laarni, J., Kojo, I. & Sams, M. 2009. The role of visual spatial attention in audiovisual speech perception. *Speech Communication*, 51, 184–193.
[8] Lavie, N. & Tsal, Y. 1994. Perceptual load as a major determinant of the locus of selection in visual-attention. *Perception & Psychophysics*, 56, 183–197.
[9] Lavie, N. 1995. Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology*, 21, 451–468.
[10] Lavie, N. 2005. Distracted and confused?: Selective attention under load. *Trends in Cognitive Sciences*, 9, 75–82.

Stylistic Effects on the Acoustic and Articulatory Properties of English Rhotics

Sarah Harper[1], Louis Goldstein[1] and Shri Narayanan[1,2]
[1]*University of Southern California Linguistics,* [2]*University of Southern California Electrical Engineering*

This study investigates acoustic and articulatory differences in the production of the English rhotic consonant across phonetic contexts and speech styles. Recent research into phonetic variability has emphasized the importance of examining production patterns in various speech styles for a more thorough view of speech behaviour. As a result of this thread of research, multiple acoustic and articulatory aspects of speech production have been observed to vary systematically across different speech styles, with measurements taken from connected speech samples often reflecting the limitations on articulatory movement imposed by the shorter duration of segments in connected or more spontaneous speech [1, 2].

The articulatory complexity of English /ɹ/, as well as the variable tongue posture observed in /ɹ/ production across different speakers and phonetic contexts [3], leads to intricate articulatory-acoustic relations that may vary in interesting ways across different speech styles. In most dialects of American English, the approximant /ɹ/ is produced by the coordination of three supralaryngeal constrictions: a labial constriction, a lingual constriction along the palatal vault, and a pharyngeal constriction in which the tongue root retracts towards the wall of the pharynx [4]. Perturbation theory predicts that, since the location of each of these constrictions corresponds to an F3 velocity maximum in a neutral tube, they will each contribute to the most salient acoustic correlate of English /ɹ/, its low F3 value [5]. As both the articulatory differences that may emerge in /ɹ/ production across different speech styles and the exact contribution of each gesture to the observed F3 lowering are largely unexamined, the results of this study shed light on these issues and expand the literature on articulatory and acoustic correlates of speech style.

Articulatory measurements of the timing and maximum constriction achieved for the three gestures involved in /ɹ/ production, as well as acoustic measurements of the first five formants, were obtained for a total of 510 tokens of English /ɹ/. Data were collected from real-time MRI recordings of three native speakers of American English (2 male, 1 female) and the accompanying audio recordings of each session. All participants were recorded reading stimuli in two speaking conditions: (1) the Citation condition (single words) and (2) the Connected Speech condition (sentences from the USC-TIMIT corpus [6]). The target words in the Citation condition all contained /ɹ/ in one of five vowel contexts and in one of four syllable contexts: a simple onset, a complex onset, a simple coda and a complex coda. Tokens in the Connected Speech condition were taken from words in the USC-TIMIT corpus that contained /ɹ/ in contexts equivalent to those in the Citation condition. For each token from both conditions, the time of the maximum constriction for each gesture was found using a region of interest technique [7] and the degree of constriction aperture was measured at this time point using an air-tissue boundary segmentation algorithm [8]. Additionally, formant values were automatically extracted at the time of maximum constriction for each gesture in each token using a script in Praat.

Our results indicate that speech style has a significant effect on the maximum degree of constriction attained by the three English /ɹ/ gestures. Interestingly, the extent of the effect on each gesture differed as a function of the token's position in the word. For onset /ɹ/ there was a greater difference observed across speech styles in the coronal constriction [Fig. 1], while in coda positions the pharyngeal constriction was more affected [Fig. 2]. Additionally, a clear relationship was found between aperture at each constriction location and the corresponding F3 measurements, with higher F3 values in the Connected Speech condition reflecting an overall lesser degree of constriction among tokens from this condition [Fig. 3].

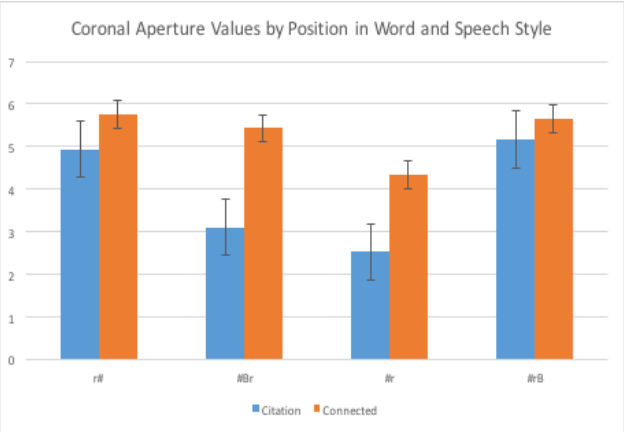


Figure 1. *Coronal aperture by position and speech style (Blue = Citation condition, Orange = Connected speech; L to R: Single coda, complex onset, single onset, complex coda)*

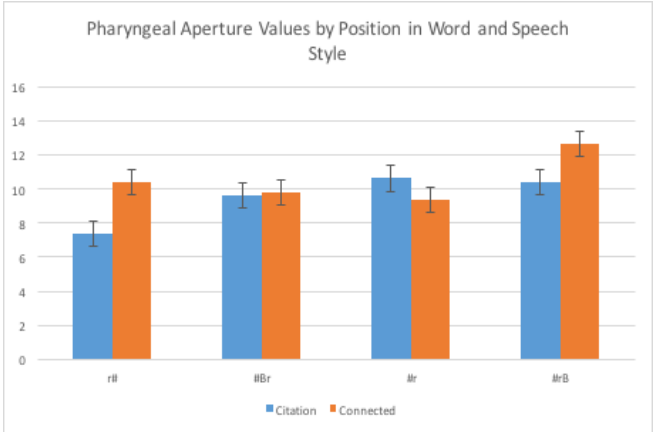


Figure 2. *Pharyngeal aperture by position and speech style (Blue = Citation condition, Orange = Connected speech; L to R: Single coda, complex onset, single onset, complex coda)*

[1] Moon, S.-J., & Lindblom, B. 1994. Interaction between duration, context and speaking style in English stressed vowels. *Journal of the Acoustical Society of America*, 96, 40-55.

[2] Lindblom, B. 1990. Explaining phonetic variation: A sketch of H&H theory. In *Speech production and speech modelling*. Springer Netherlands, 403-439.

[3] Guenther, F., Espy-Wilson, C., Boyce, S., Matthies, M., Zandipour, M, & Perkell, J. 1999. Articulatory trade-offs reduce acoustic variability in American English /r/ production. *Journal of the Acoustical Society of America*, 105, 2854-2965.

[4] Alwan, A., Narayanan, S., & Haker, K. 1997. Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part II: The rhotics. *Journal of the Acoustical Society of America*, 101, 1078-1089.

[5] Espy-Wilson, C., Boyce, S., Jackson, M., Narayanan, S., and Alwan, A. 2000. Acoustic modeling of American English /r/. *Journal of the Acoustical Society of America,* 108, 343-356.

[6] Narayanan, S., Toutios, A., Ramanarayanan, V., Lammert, A., Kim, J., Lee, S., … and Proctor, M. 2014. Real-time magnetic resonance imaging and electromagnetic articulography database for speech production research (TC). *Journal of the Acoustical Society of America*, 136(3), 1307–1311.

[7] Tiede, M. 2010. MVIEW: Multi-channel visualization application for displaying dynamic sensor movements. Computer program.

[8] Kim, J., Kumar, N., Lee, S., & Narayanan, S. 2014. Enhanced airway-tissue boundary segmentation for real-time magnetic resonance imaging data. *ISSP 10* (Cologne), 222-225.

Czech ToBI

Andrea Pešková
University of Osnabrück

Whereas many languages (e.g., Catalan, English, German or Greek) already come with a long tradition of analysis from within Autosegmental Metrical (AM) model of intonational phonology ([1]; [2]), intonational properties of Czech have been studied very little, or examined outside any theoretical framework (for an exception see [3] who offers preliminary descriptions within the AM model). The aim of this paper is to fill this gap and to present a first proposal for a Czech Tone and Break Indices (ToBI) labelling system, grounded in the AM model. As for some prosodic properties of this West Slavic language, Czech is an intonational language, it has a fixed stress on the first syllable, and it is considered as a kind of mix-timed language for its rhythmic features (see, e.g. [4]). For the main purpose of the present paper (the annotation of the intonation of spoken utterances), the data were elicited by means of a questionnaire based on the so-called Discourse Completion Task (see, e.g., [5]), as well as, on the supplementary reading tasks. 10 speakers of the Southern Moravian dialect were recorded, and a total of 300 sentences were obtained. The corpus comprises intonational patterns of different types of sentences such as statements, yes-no questions, wh-questions, imperatives, and vocatives (in addition, the sentences had also a specific pragmatic meaning).

First results show that the Czech variety under discussion has (at least) eight different nuclear configurations (Table 1; Fig. 1–8). It thus displays a smaller intonational inventory in comparison with some other languages like, for example, Catalan or Spanish, which have fifteen and nineteen different nuclear configurations, respectively ([6]; [7]). Comparing it with these two Romance varieties, Czech presents only three boundary tones: L%, LH%, !H%. Moreover, the downstepped high boundary tone !H has also an additional phonetic realization: a long duration associated with the last syllable of the utterance. It will be argued that this lengthening property, found in insistent and exhort requests, represents a phonological contrast and therefore a preliminary label !H:% is proposed here. As for prenuclear pitch accents, the typical realization is a low tone on a stressed syllable, followed by a rise (L*+H). A further interesting observation is that Czech solves the tonal crowding (i.e. a context where two or more tones are associated with the same segmental element) by means of tonal compression. There remains a question if and in which way Czech uses different syntactic strategies (e.g., word order), lexical markers and/or modal particles which may impact on pragmatic meaning of the utterances instead of intonation, as it was stated, for example, in Friulian (see [8]).

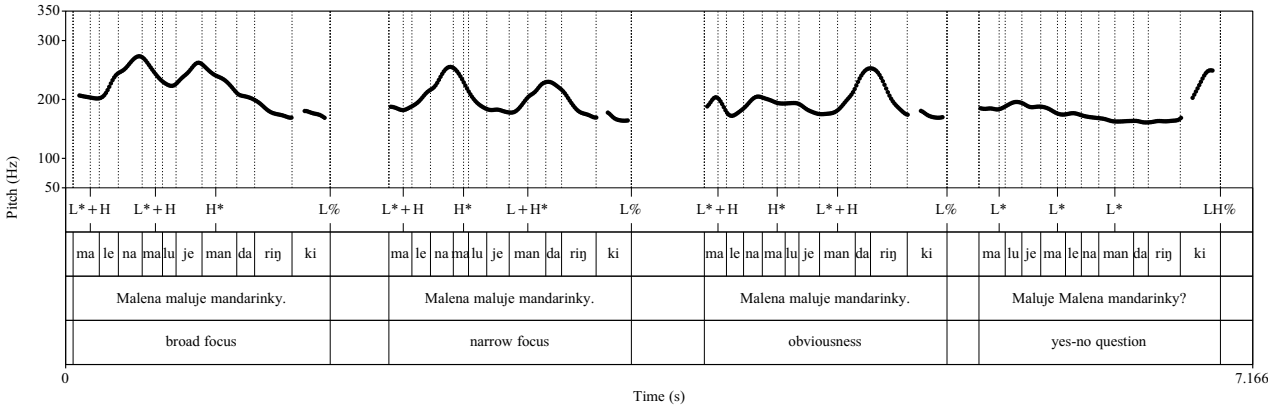
The research hopes to contribute not only to the study of intonational properties in Czech, and cross-linguistic variation, but also to the study of intonation grammar in general.

[1] Pierrehumbert, J. B. 1980. *The Phonology and Phonetics of English Intonation*. Cambridge, MA: MIT Press.
[2] Gussenhoven, C. 2004. *The Phonology of Tone and Intonation*. Cambridge: CUP.
[3] Duběda, T. 2011. *Towards an inventory of pitch accent for read Czech*. Slovo a slovesnost 72, 3–11.
[4] Dankovičová, J. & V. Dellwo. 2007. *Czech Speech Rhythm and the Rhythm Class Hypothesis*. Proceedings of the 16th ICPhS.
[5] Frota, S. & P. Prieto (eds.). 2015. *Intonational Variation in Romance.* Oxford: OUP.
[6] Prieto P., et al. 2015. *Intonational phonology of Catalan and its dialectal varieties*. In: *Intonational Variation in Romance* (eds. S. Frota & P. Prieto), 9–62.
[7] Aguilar, L., et al. 2009. *Sp_ToBI Training Materials*. http://prosodia.upf. edu/sp_tobi/.
[8] Roseano P., et al. 2015. *Intonational phonology of Friulian and its dialects*. In: *Intonational Variation in Romance* (eds. S. Frota & P. Prieto), 101–139.

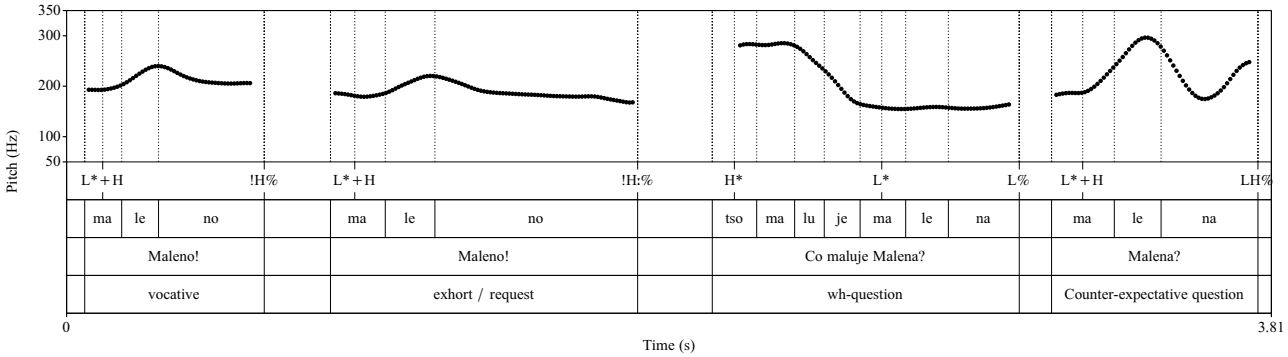
Appendix

Nuclear configuration	Type of sentence
L* LH%	Neutral yes-no question. Echo yes-no question. Confirmatory yes-no question. Imperative yes-no question. Polite yes-no question.
L*+H LH%	Counter-expectative yes-no question.
H* L%	Broad focus statement. Exhort / request (neutral). Order / imperative. Exclamative wh-question. Exhortative wh-question.
L+H* L%	Narrow focus statement (Fc = object).
L*+H L%	Emphatic exclamative. Obviousness statement. Order / imperative. Hesitation statement. Exhort / request (neutral).
L* L%	Neutral wh-question. Imperative wh-question.
L*+H !H%	Neutral vocative. Insistent vocative.
L*+H !H:%	Exhort / insistent request (familiar).

Table 1: *Nuclear configurations in different type of sentences.*



Figures 1-4 (from left): *Intonation contours of different sentence types I.*



Figures 5-8 (from left): *Intonation contours of different sentence types II.*

Translation of the sentences (Figures 1–8):
Malena maluje mandarinky. ‘Malena draws tangerines.’
Maluje Malena mandarinky? ‘Does Malena draw tangerines? lit. Draws Malena tangerines?’
Maleno!(*?*) ‘Malena!(?)’
Co maluje Malena? ‘What does Malena draw? lit. What draws Malena?’

Monday, 13:15 - 14:15
Poster Session 1, #3

Perceptual similarity spaces of British English vowels by speakers of Pakistani Urdu

Ishrat Rehman & Amalia Arvaniti
University of Kent

An important issue in L2 acquisition is how learners assimilate the sounds of the L2 into their L1 system [1, 2]. Understanding this process provides us not only with insights about L2 acquisition, but also about the relative weight of phonetic dimensions for speakers of different languages; such differences in relative weight are of interest as they can affect percepts of similarity (cf. [3]). Understanding these effects can be achieved using *free classification*, in which listeners group L2 segments based on perceived similarity but without being given predetermined categories from their L1 with which to match the L2 segments [4]. This is undertaken here with Pakistani Urdu speakers; these speakers use English regularly and have developed a distinct variety of English (Pakistani English) [5], [6], [7]. The present work serves to shed light on the features of English vowels that have been most relevant in developing and sustaining this "new English" variety.

Seventy listeners (46 F, 24 M) from Lahore, Punjab, Pakistan took part in the experiment. They were University students between 18 and 24 years of age with Punjabi and Urdu as their first two languages (Pakistanis learn Urdu, a national language of Pakistan, at school but typically speak a regional language, like Punjabi, at home). The experiment was conducted at the University of the Punjab in Lahore. The stimuli were 19 *hVd* words carrying the Standard Southern British English (SSBE) vowels /iː ɪ e æ ɜː ʌ ɑː ɒ ɔː ʊ uː ɪə eə ʊə eɪ aɪ ɔɪ əʊ aʊ/ produced by two native speakers of that variety (1F, 1M). The stimuli were presented to listeners on the left hand side of a PowerPoint slide in three columns of 19 randomly coded boxes, each containing a sound file with one of the *hVd* words. The participants were free to listen to the stimuli as many times as they wanted; their task was to group the words with similar sounding vowels by dragging the boxes onto a grid on the right hand side of the slide. Most participants took 25-55 minutes to complete the task. The responses were converted into similarity matrices and statistically analysed using hierarchical clustering and multidimensional scaling [8]. These two analyses, which presented similar patterns, allowed us to examine the participants' overall classification strategy and visualise the perceptual similarity spaces.

For monophthongs, listeners were sensitive to both F1 and F2, but found it difficult to distinguish the mid-high from the mid-low vowels of English, both front and back (Figures 1a, 2a). The SSBE central vowels /ɜː/ and /ʌ/ were not grouped together, indicating that the participants lack a central vowel space; /ɜː/, which in SSBE is somewhat more fronted than /ʌ/ [9], was grouped with the front vowels, and /ʌ/ were grouped with the back vowels. This suggests that the listeners prioritized F2 relative to F1. Further, duration was less important than quality, in that listeners grouped /iː/ and /ɪ/, /ʊ/ and /uː/ or /ɔː/ and /ɒ/ with each other. The diphthongs were often grouped with monophthongs, though not in a consistent manner: sometimes the groupings were based on the initial and sometimes on the final element (Figures 1b, 2b). This did not apply to /aʊ/ /aɪ/ and /ɔɪ/, which were grouped together, possibly because they are the SSBE diphthongs in which the first and last element are most distinct; they were possibly the SSBE diphthongs perceived as such by the Punjabi-Urdu listeners.

Overall, the results show that Punjabi-Urdu listeners are sensitive to both vowel height and backness, but possibly prioritize the latter over the former, while they lack an intermediate (central) space. This last observation is further supported by their groupings of diphthongs whose elements are least distinct from each other and which they grouped with monophthongs. Overall these results give credence to the idea of cue weighting for vowels. Further, our conclusions are consistent with features of Pakistani English described in [5], [6] and [7]; e.g. the fact that the partricipants group diphthongs like /eɪ/ and /əʊ/ with monophthongs may be the reason why these diphthongs are absent from Pakistani English.

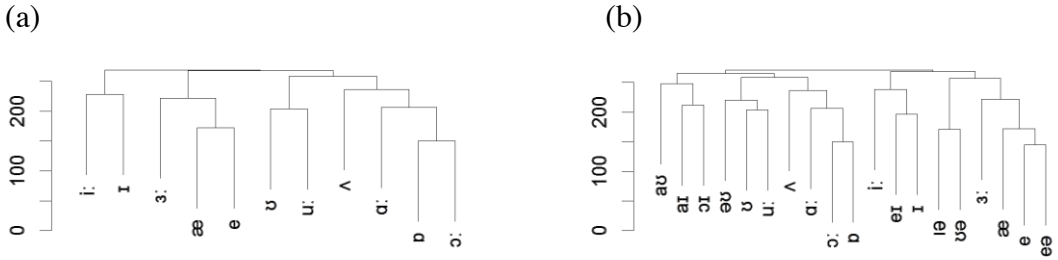


Figure 1. *Hierarchical clustering of monophtongs (a) and monophthongs and diphthongs (b).*

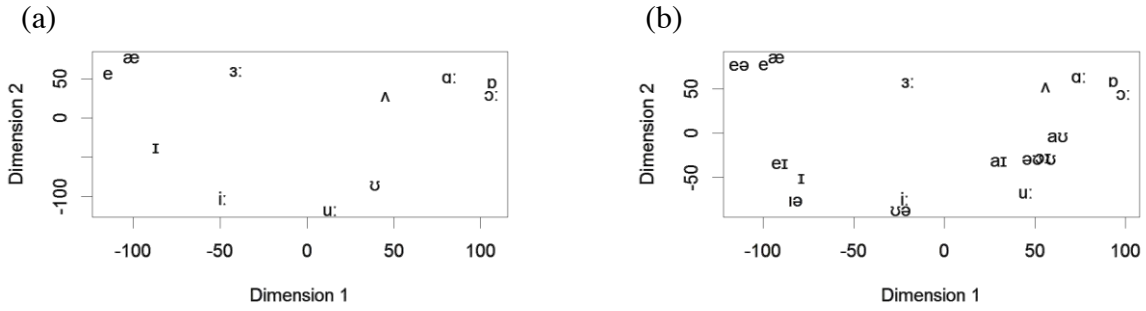


Figure 2. *Two-dimensional MDS of perceptual similarity for monophthongs (a) and monophthongs and diphthongs (b).*

[1] Best, C.T. & Tyler, M.D., 2007. Nonnative and second-language speech perception: Commonalities and complementarities. In M. J. Munro & O.-S. Bohn (eds), *Second language speech learning: The role of language experience in speech perception and production*, pp. 13-34. Amsterdam: John Benjamins.

[2] Flege, J.E., Bohn, O.S. & Jang, S., 1997. Effects of experience on non-native speakers' production and perception of English vowels. *Journal of phonetics* 25(4): 437-470.

[3] Hwang, H. K. 2016. A crosslinguistic Study of Vowel Categorization: Data from Canadian English, Korean and Japanese. Poster presented at 15th Conference on Laboratory Phonology, Cornell University, 13-17 July 2016.

[4] Clopper, C.G. 2008. Auditory free classification: Methods and analysis. *Behavior Research Methods* 40(2): 575-581.

[5] Sheikh, Q.A., 2012. An analysis of the vowel sounds of Pakistani English. *Bulletin of Education and Research 34*(1): 1-18.

[6] Platt, J. T., Weber, H., & Ho, M. L. (1984). *The New Englishes*. London: Routledge & Kegan Paul.

[7] Rahman, T. (1991). Pakistani English: Some Phonological and Phonetic Features. *World Englishes,* 10(1), 83-95.

[8] Venables, W. N. & Ripley, B. D. (2002). *Modern Applied Statistics with S*. Fourth Edition. Springer: New York.

[9] Hawkins, S. & J. Midgley. 2005. Formant frequencies of RP monophthongs in four age groups of speakers. *Journal of the International Phonetic Association* 35(2): 183-199.

Monday, 13:15 - 14:15
Poster Session 1, #4

Vowel spaces in six regional Greek varieties: an acoustic and articulatory analysis

Mary Baltazani[1], Katerina Nicolaidis[2] and Anastasia Chionidou[2]
[1]University of Oxford, [2]Aristotle University of Thessaloniki

Vocalic system surveys shed light on the principles behind their organization. The same 5-7 vowels occur in 75% of vowel inventories [1]. Such proclivities have given rise to hypotheses on the forces shaping vowel spaces: Dispersion Theory postulates that the distance among vowels serves to reduce confusion and increase perceptual contrast [2], [3]; the Quantal Theory of speech [4] suggests there are regions of stability in the phonetic space, in particular for /i/, /a/, /u/, regardless of vocalic inventory size. We present evidence suggesting a more complex picture than either of these hypotheses assume.

Acoustic and articulatory (EPG) analyses martial vowel data from 6 regional varieties of Greek (Athens, Crete, Epirus, Macedonia, Peloponnese and Thessaly), most of which are endangered. The vowels were produced by 12 speakers, 6F, 6M, balanced for dialect, in two read speech tasks, where V=/i, e, a, o, u/: (a) in symmetric /pVpV/ words in two stress conditions (1st or 2nd syllable); (b) in /C^1VC^2a/ words, $C^{1,2}$=[labial, alveolar, palatal, velar], in two stress conditions, resulting in 4200 tokens. Acoustic measurements include the duration, F1XF2 acoustic space and formant distances between adjacent vowels, while articulatory ones include the Centre of Gravity and the mean lateral measures for the examination of the place and degree of vocalic constriction.

Preliminary results indicate an effect of all the above factors, with dialect and stress exerting the strongest influence for the observed variability. Northern dialects (Epirus, Macedonia, Thessaly) are characterised by two phonological processes, unstressed high vowel deletion and unstressed mid vowel raising, which contribute towards differentiating northern from southern vocalic spaces (Fig. 1, left and right respectively). Also, while the five Greek vowels occupy roughly the same position in the F1XF2 space across the three southern dialects examined, this is not the case for northern dialects, where vowels are not maximally dispersed. One of the most remarkable findings is the minimization of the /i/ - /e/ distance in the northern dialects, especially in Epirus Greek where the two totally overlap. This finding is also verified articulatorily: the lateral and very front contact patterns are similar for the unstressed /i/ and /e/ (Fig. 2, top left/right), and for the stressed /i/ (Fig. 2, bottom left), but not for the stressed /e/ (Fig. 2, bottom right) which does not show much contact in these areas.

These intriguing findings suggest that the explanation behind vocalic space structure is complex and the variation observed may relate to several factors. Previous literature indicates that acoustic vowel targets can differ in languages/dialects with the same number of vowel contrasts due to phonological or historical reasons; thus frequency intervals between adjacent vowels can relate to dialect-dependent patterns rather than to universal ones [5], [6]. Such patterns are seen in the Greek dialects. For example, northern dialects are known for massive vowel deletions [7] which lead to more closed syllables and a greater degree of palatalization in their consonantal inventory, factors which correlate with the observed shift in vocalic space (cf. [5]). Differences may also occur due to context-dependent variability across dialects, such as consonantal inventories and inherent vowel durations in different rates of production.

Overall, our work showcases how indispensable cross-dialectal variation studies are for the deeper understanding of speech production variation. We will discuss dialect-specific sources of variability under the scope of current theories on i) vowel dispersion, ii) direction of unstressed phonetic reduction, and iii) sociolinguistic variation across genders.

[1] Maddieson, I. 1984. *Patterns of sound*. Cambridge University Press.
[2] Liljencrants, J., Lindblom, B. 1972. Numerical Simulation of Vowel Quality Systems: The Role of Perceptual Contrast. *Language* 48, 4: 839-862.
[3] Lindblom, B. 1986. Phonetic universals in vowel systems. In: Ohala JJ, Jeager JJ (Eds.) *Experimental Phonology*. Orlando, FL: Academic Press: 13-44.
[4] Stevens, K.N. 1989. On the Quantal nature of speech. *Journal of Phonetics* 17: 3-45.
[5] Bradlow, A.R. 1995. A comparative acoustic study of English and Spanish vowels. *JASA*, 97(3): 1916-1924.
[6] Recasens D., Espinosa, A. 2006. Dispersion and variability of Catalan vowels, *Speech Communication*, 48 : 645-666.
[7] Topintzi N. & Baltazani, M. 2012. The acoustics of high-vowel loss in a Northern Greek dialect and typological implications. In *Consonant Clusters and Structural Complexity*, Hoole, P. et al. (Eds.), de Gruyter: 373-402.

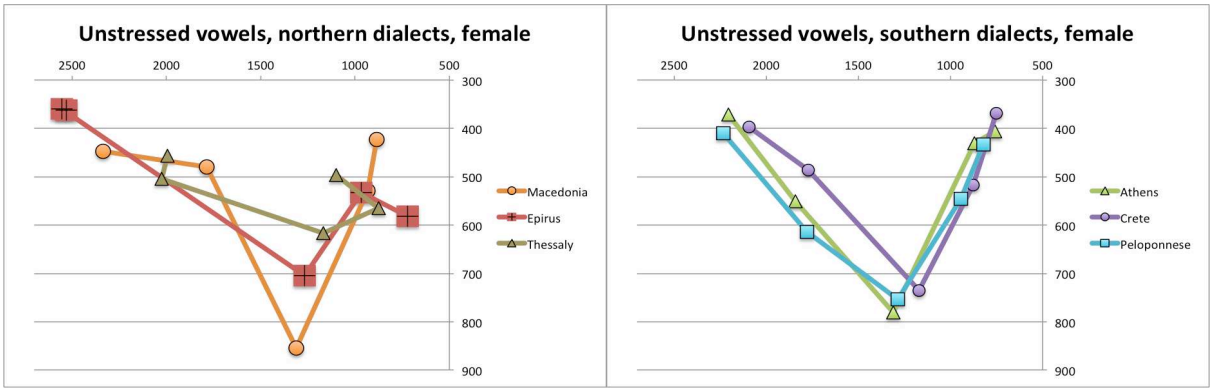


Figure 1: *Position of unstressed vowels in the F1XF2 space in northern (left) and southern (right) Greek dialects, as produced in /p_p/ context by female speakers.*

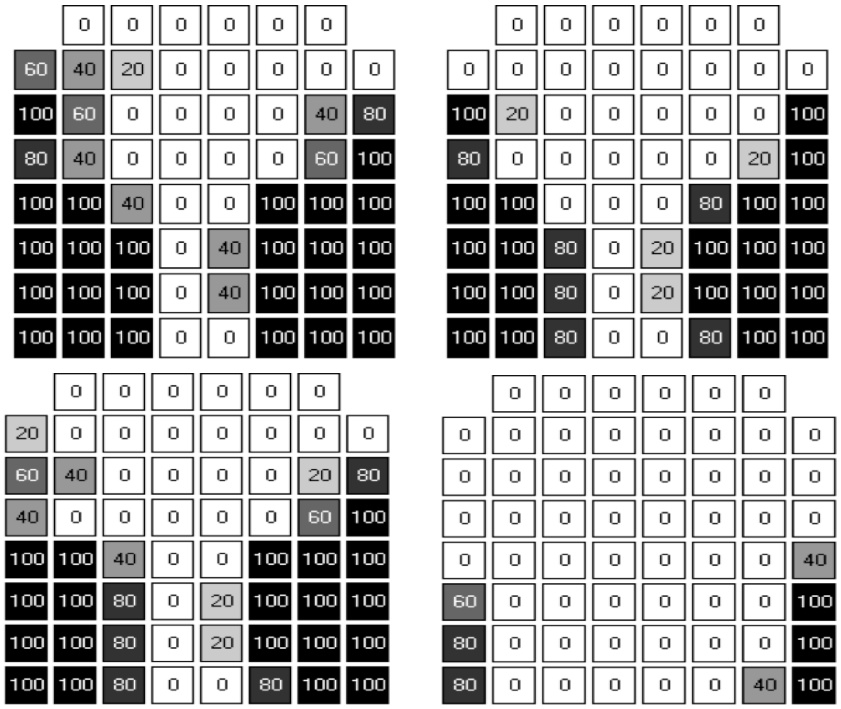


Figure 2: *Palatogram of **unstressed** /i/ and /e/ (top left/right) and **stressed** /i/ and /e/ (bottom left/right) in /p_p/ context, female Epirus speaker*

Flexibility of the acoustics-to-articulation mapping: evidence from a bidirectional perturbation study

Eugen Klein[1], Jana Brunner[1], Phil Hoole[2]
[1]Humboldt University of Berlin, [2]Ludwig Maximilian University of Munich

A recent experiment combining auditory and articulatory perturbation has shown that participants did not compensate for extreme articulatory perturbations as long as the produced auditory signal was altered to match their desired output (Feng, Gracco, and Max, 2011). These findings suggest that participants monitor their speech production in terms of acoustic targets as they seem to accept a modified articulatory configuration as long as the acoustic output conforms to their production intention. This result predicts high flexibility of the acoustics-to-articulation (A-to-A) mapping in terms of its rapid adaptation mechanism which would allow participants to produce an intended speech sound with different articulatory configurations. We present results from a bidirectional perturbation study which assessed this prediction along with the question of how feasible it is for a speaker to maintain a one-to-many relation between acoustic and articulatory targets.

Speech of 18 native Russian speakers was acoustically recorded while they were producing CV syllables which contained the central unrounded vowel /ɨ/. The second formant (F2) of the vowel /ɨ/ was perturbed in real-time during the production and fed back to the participant via insert headphones. The F2 was shifted upwards on one half of the trials and downwards on the other half in order to encourage participants to produce the vowel /ɨ/ with two different articulatory configurations. The trials were presented in random order. The direction of the perturbation was dependent on the consonant preceding /ɨ/ and was counterbalanced for the place of articulation between the 18 participants to assess the influence of the preceding consonant on the amount of compensation in the vowel /ɨ/. Each experimental session contained a baseline phase (without perturbation) followed by three perturbation phases which differed in the amount of the applied perturbation between 220, 370, and 520 Hz.

Our analyses show that, within the same experimental session, participants produced significantly different F2 values for the different tokens of /ɨ/ depending on the direction and the amount of applied perturbation. The upper panel of the Figure 1B shows F2 values produced by nine speakers when F2 was perturbed with 520 Hz downwards for the syllable /dɨ/ and upwards for /gɨ/. In line with our expectations, the produced F2 values drifted apart for seven participants in this group. On that, compare the difference in F2 between /dɨ/ and /gɨ/ in the upper panel of the Figure 1A (baseline phase) with the difference in the upper panel of the Figure 1B (last perturbation phase). The upper panel of the Figure 1C shows F2 data during the last perturbation phase for the remaining nine speakers, in whose production, in contrast to the first group, F2 was perturbed upwards for the syllable /dɨ/ and downwards for the syllable /gɨ/. In line with our expectations, the produced F2 values for /dɨ/ and /gɨ/ crossed with increasing amount of perturbation for all participants of the second group. Nonetheless, there were inter- and intra-individual differences. Most of the participants, for instance, additionally changed their F3 values in the syllable /dɨ/ which suggests that in this particular context, aside of the horizontal tongue position, they were using lip rounding to compensate for F2 shifts. On that, compare bottom panels of the Figures 1A, 1B, and 1C.

In summary, our findings based on the bidirectional perturbation design provide evidence for the hypothesis that A-to-A mapping is highly flexible and allows for a one-to-many relation between acoustic and articulatory targets.

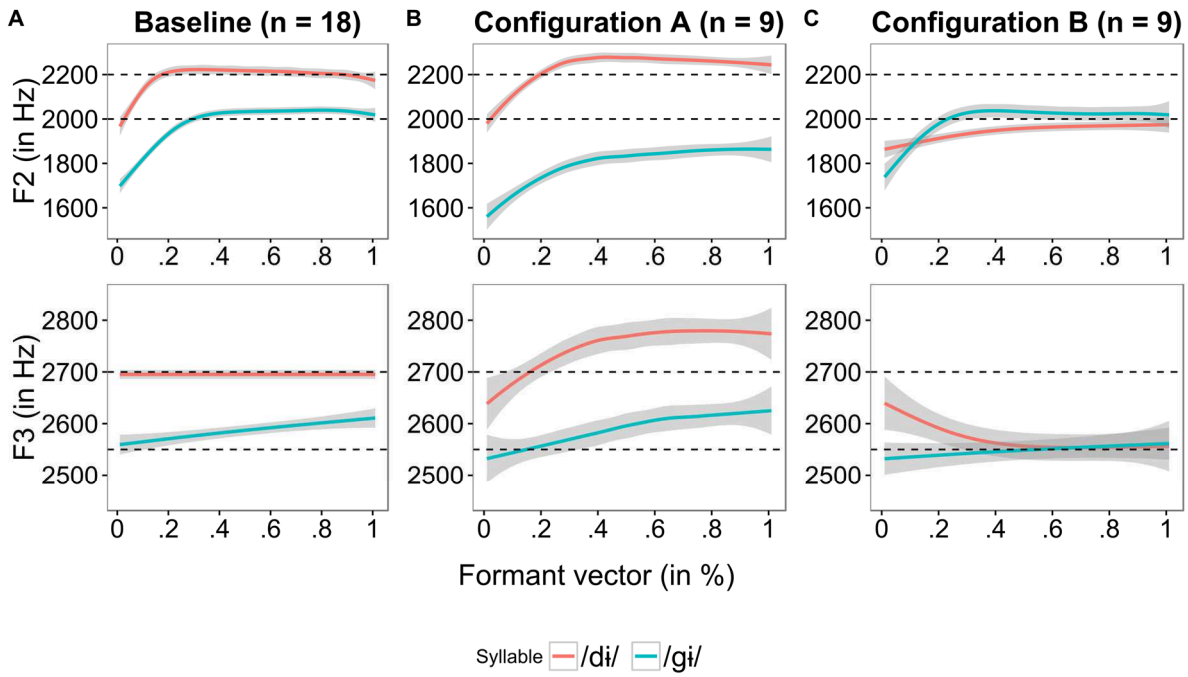


Figure 1. *F2 (upper panels) and F3 (bottom panels) formant vectors produced by the participants during the baseline (1A) and the last perturbation phase (1B and 1C). The data of the last perturbation phase is plotted separately for two experimental groups based on their perturbational configuration (1B and 1C).*

[1] Feng, Y., Gracco, V.L., Max, L. 2011. Integration of auditory and somatosensory error signals in the neural control of speech movements. *J. Neurophysiol.*, 106, 667– 679.

Articulatory overlap in a subset of stop+lateral clusters in Spanish

M. Gibson[1], S. Sotiropoulou[2], S. Tobin[2] & A. Gafos[2,3]
[1]Universidad de Navarra, [2]Universität Potsdam, [3]Haskins Laboratories

Place order effects (POE) express a generalization found for a number of languages toward more gestural overlap in certain sequences based on the anterior/posterior ordering of the different articulators [1-7]. Sequences which start with an anterior articulator (lips, for example) and progress backward toward the posterior region have been found to exhibit more gestural overlap than sequences which start with a posterior articulator (tongue back, for example) and move forward to a more anterior articulator, the former being known as a front-to-back (FtB) while the latter being referred to as a back-to-front (BtF) order.

The current study addresses the temporal organization of gestures in a subset of word-initial stop-liquid (/kl, gl, pl, bl/) clusters in Spanish. In addition to articulatory overlap, the duration of the gestures is also examined, as duration of target plateaux has been shown to relate to syllable margins [8-17].

Data were collected using an electromagnetic articulograph (EMA, Carstens AG501) from three native speakers of Central Peninsular Spanish at the Phonetics Laboratory of the Universität Potsdam. Overlap in C1C2 clusters was defined as per [7, 9] by the inter-plateau interval (IPI) between the two C constrictions, while duration was calculated by subtracting the plateau onset timestamp from the plateau release timestamp (d-c/b-a = C duration, see Figure 1 below).

Mview, a computer program designed by Mark Tiede, was used to label the kinematic files and extract variable values. Dependent variables (IPI as defined above, C1 and C2 duration) were subject to log transformations before analysis. PLACE ORDER (BtF, FtB) and CLUSTER VOICE (uniform voice and mixed voice clusters) were modeled as predictors. A series of linear mixed effects models using the lme4 package in R [18] were employed for the statistical analyses. Speaker and repetition were modeled as random effects. For post-hoc comparisons, significance was determined using the Tukey adjusted contrast.

Pooled means show a main effect ($\chi^2[1, N=535] = 534$, $p=0.02$) for place order in the expected direction, front-to-back sequences yielding shorter temporal transitions, as measured by IPI, between articulatory plateaux than back-to-front clusters. Voicing, however, showed a stronger effect ($\chi2[1, N=535] = 534$, $p<0.001$) than place on IPI patterns, uniform voice clusters showing lower positive lag times between the plateaux than mixed voice clusters (see Table 1). Nevertheless and in line with French [9], in either voicing condition (uniform or mixed), Spanish exhibits robust open transitions between the two gestures in stop-liquid clusters.

As for the effects of voice and place on the duration of the target plateaux of C1 and C2 (lateral), main effects were found for both voice ($\chi2[1, N=535] = 534$, $p< 0.001$) and place ($\chi2[1, N=535] = 534$, $p=0.03$) on C1 (see Table 2), as well as their interaction ($\chi2[1, N=535] = 534$, $p< 0.001$). As for lateral duration, both place ($\chi2[1, N=535] = 534$, $p< 0.001$), and voice ($\chi2[1, N=535] = 534$, $p< 0.001$) specifications of the preceding C1 were found to have a strong main effect on the temporal duration of C2 (see Table 3), laterals appearing after velars being longer than after labials, and laterals following voiced consonants being longer than following voiceless gestures.

The results offer a first report of articulatory overlap in Spanish onsets and make clear the articulatory basis for the vocoid-like element between gestures in CC clusters which have before been documented in acoustic data [e.g., see 19 among others].

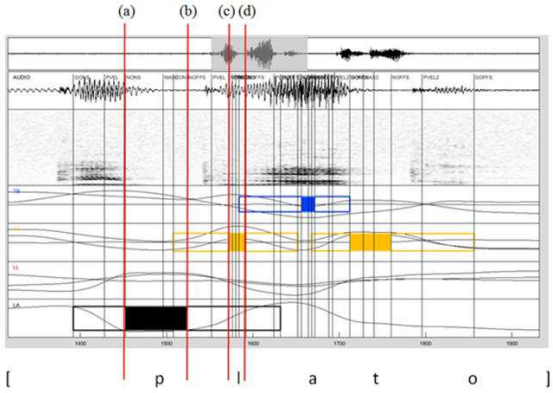


Figure 1. Parses for the individual gestures (as consonants are in principle sets of gestures, e.g. oral and laryngeal) in the token *plato* 'plate'. The (a) timestamp (NONS, on the Lip Aperture signal) marks the onset of the target articulatory plateau for /p/, (b) marks the offset of the articulatory target plateau of /p/ (NOFFS, on the Lip Aperture signal), (c) marks the onset of the tongue-tip target for /l/ (NONS, on the TT signal), while (d) marks the offset of the tongue-tip target for /l/ (NOFFS, on the TT signal).

Table 1. Means and standard deviation of interplateau intervals as a function of voice and place.

nonzero interplateau interval (ms)		
	Means	*sd*
Mixed-voice	-35.1	28.7
Labial	-37.5	28.5
Velar	-32.9	29.1
Uniform-voice	-25.8	23.7
Labial	-24	19.3
Velar	-27.6	28.1

Table 2. Means and standard deviations for [C1] duration.

	Duration C1 (ms)			
	Labial	*sd*	Velar	*Sd*
Voice	35.2	27.9	30.4	21.8
Voiceless	45.5	23.4	41.1	23.4

Table 3. Means and standard deviations by factor for lateral duration (ms)

	Duration C2 ([l], ms)			
	Labial	*Sd*	Velar	*sd*
Voice	35.2	23.2	40.4	23.0
Voiceless	27.6	18.8.	29.4	18.1

References

[1] Hardcastle, William and Peter Roach. 1979. An instrumental investigation of coarticulation in stop consonant sequences. In: Harry Hollien and Patricia Ann Hollien (eds.), *Current Issues in the Phonetic Sciences*, 531–540. Amsterdam: John Benjamins.
[2] Byrd, Dani. 1992. Perception of assimilation in consonant clusters: A gestural model. *Phonetica* 49: 1–24.
[3] Byrd, Dani. 1996. Influences on articulatory timing in consonant sequences. *Journal of Phonetics* 24: 209–244.
[4] Zsiga, Elisabeth C. 1994. Acoustic evidence for gestural overlap in consonant sequences. *Journal of Phonetics* 22: 121–140.
[5] Surprenant, Aimée M. and Louis Goldstein. 1998. The perception of speech gestures. *Journal of the Acoustical Society of America* 104(1), 518–529.
[6] Chitoran, Ioana, Louis Goldstein and Dani Byrd. 2002. Gestural overlap and recoverability: Articulatory evidence from Georgian. In: Carlos Gussenhoven, Toni Rietveld and Natasha Warner (eds.), *Laboratory Phonology VII*, 419–447. Berlin/New York: Mouton de Gruyter.
[7] Gafos, A., Hoole, P., Roon, K., Zeroual, C. 2010. Variation in timing and phonological grammar in Moroccan Arabic clusters. Laboratory Phonology 10: Variation, Detail and Representation, Ed. Cécile Fougeron (Mouton de Gruyter: Berlin/New York).
[8] Haggard, M. 1973. Abbreviation of consonants in English pre-post-vocalic clusters, *Journal of Phonetics*,3, 7–24.
[9] Bombien, L. and P. Hoole. 2013. Articulatory overlap as a function of voicing in French and German consonant clusters, *Journal of the Acoustical Society of America*, volume 134(1), 539-550
[10] O'Shaughnessy, D. 1974. Consonant durations in clusters, *IEEE Transactions on Acoustics Speech and Signal Processing*, 22, 282-295.
[11] Marin, S. and L.S. Bučar Shigemori. 2014. Vowel Compensatory Shortening in Romanian, [online], http://phonetik.unimuenchen.de/universals/pub/ISSP2014_VCS_Marin_BucarShigemori.pdf
[12] Lisker, L. and A.S. Abramson. 1967. Some Effects of Context on Voice Onset Time in English Stops, *Language and Speech*, 10, 1–28.
[13] Menon, K.M.N, P.J. Jensen and D. Dew. 1969. Acoustic properties of certain VCC utterances, *Journal of Acoustical Society of America*, 46, 449-457.
[14] Suen, C. Y. and M. P. Beddoes. 1974. The silent interval of stop consonants, *Language and Speech*, 17, 126-134.
[15] Bailey, P.J. and Q. Summerfield. 1980. Information in speech: Observations on the perception of [s]-stop clusters, *Journal of Experimental Psychology: Human Perception and Performance*, 6, 536–563.
[16] Stathopoulos, E.T. and G. Weismer. 1983. Closure duration of stop consonants, *Journal of Phonetics*, 11, 395–400.
[17] Gibson, M., Fernández Planas, A.M., Gafos, A., and Ramirez, E. 2015. Consonant duration and VOT as a function of syllable complexity and voicing in a sub-set of Spanish clusters. Proceedings of Interspeech 2015. Dresden, Germany. 1690-1694.
[18] R Core Team, "R: A language and environment for statistical computing", R Foundation for Statistical Computing, Vienna, Austria, 2013, ISBN 3-900051-07-0, [online] URL http://www.R-project.org/.
[19] Colantoni, L. and J. Steele. 2005. Liquid Asymmetries in French and Spanish. In Fregeni C., Hirayama, M., and Mackenzie, S. (Eds.). Toronto Working Papers in Linguistics. Special Issue on Similarity in Phonology. Toronto. Univ. of Toronto.

Museum or _musical_? – Pitch accent type affects word recognition in Australian English

Katharina Zahner[1], Heather Kember[2], Anne Cutler[2], Bettina Braun[1]
[1]*University of Konstanz,* [2]*The MARCS Institute for Brain, Behaviour & Development*

Intonation languages signal post-lexical information (e.g. information structure) by means of different pitch accent types (e.g. [1]). Phonetically, high and low tonal targets may appear before, on, or after a stressed syllable and f0 is hence not a *reliable* cue to lexical stress. Recent studies on German suggest that (phrase-level) *pitch accent type* affects (word-level) stress perception and lexical access [2, 3]: German adults temporarily fixated SWW-words (e.g. *Libero*) when they heard WSW-words (e.g. *Libelle*) with an early-peak accent (H+L*, H+!H*), i.e. words with a high-pitched but unstressed initial syllable [3]. We here test whether such an effect would also appear in English – a language in which stress is not only cued by suprasegmentals (cf. duration, intensity), but by vowel reduction in many unstressed syllables [4]. Thus, English listeners make less use of suprasegmental stress cues in word recognition than Dutch and German listeners do [5-9], but if forced to use suprasegmentals they primarily rely on pitch information [10]. Using the visual world paradigm with 4 printed words on screen (cf. [11]), we thus test whether English listeners also immediately activate SWW-words (*musical*) if presented with WSW-words (*museum*) with early peaks (H+!H*).

As in [3], we selected 64 frequency-matched cohort pairs that were segmentally identical until at least the first consonant of the second syllable, but differed in the position of stress (32 were disyllabic: WS vs. SW (*cartoon* - *carton*) and 32 trisyllabic: WSW vs. SWW (*museum* - *musical*)). Note that the first syllable in a WS(W)-word always contained a full (non-reduced) vowel. Each pair was combined with 2 unrelated distractors to be presented on screen. For 32 of the 64 cohort pairs, the auditory target ("*The next word is TARGET*") was one of the cohort members (in 16 experimental trials the WS(W)-word, in 16 distractor trials the SW(W)-word); in 32 filler trials it was one of the unrelated distractors. Crucially, in experimental trials, the WS(W)-target was either produced with a medial-peak accent (L+H*) or an early-peak accent (H+!H*), manipulated in a Latin-Square Design. Targets were spliced into a semantically non-constraining carrier sentence and acoustically matched for syllable duration and f0-excursion across intonation conditions. Forty Australian English (AusE) participants (∅=25.7 years, SD=7.5 years, 22 female) were tested, using SR Eyelink1000.

Fixations to the 4 words on screen were extracted in 4ms steps (see Fig. 1). Empirical logits of fixations for experimental trials were analysed, modelling *intonation condition* as fixed and *participants* and *items* as crossed random factors in a *lmer*. During the processing of the segmentally ambiguous part, participants fixated the SW(W)-competitor (*musical*) more when the WS(W)-target (*museum*) was produced with an early peak, H+!H*, (elogs: -1.52) than when produced with a medial peak, L+H*, (elogs: -2.05; ß=0.5 [0.01;1.04], SE=0.26, t=2.02, $p<0.05$). This tendency was preserved even after segmental disambiguation towards the end of the target (ß=0.6 [0.13;1.13], SE=0.25, t=2.49, $p<0.05$, see Fig. 2). Thus, AusE listeners perceive high-pitched but unstressed syllables temporarily as stressed.

Our findings provide evidence that AusE listeners can interpret high pitch as a cue to word stress if segmental information is ambiguous (see also [5,10]), although pitch accent type differences make high pitch in fact unreliable for this purpose. The acoustic salience of high-pitched syllables, and their frequent occurrence in AusE [12], may of course help make their use more likely. Note though that the target and competitor activation seen here in AusE occurred later in time than in German [3], an asymmetry which we interpret as indicating that AusE listeners experience higher processing costs for this use of suprasegmental stress cues. In future experiments, we will test whether competitor activation is affected by the frequency of occurrence of high-pitched stressed syllables in the input, and whether other accent types with low-pitched stressed syllables lead to comparable activation patterns.

Figure 1. *Evolution of fixations to words on screen in two intonation conditions ((a) medial-peak and (b) early-peak condition); boundaries refer to acoustic landmarks.*

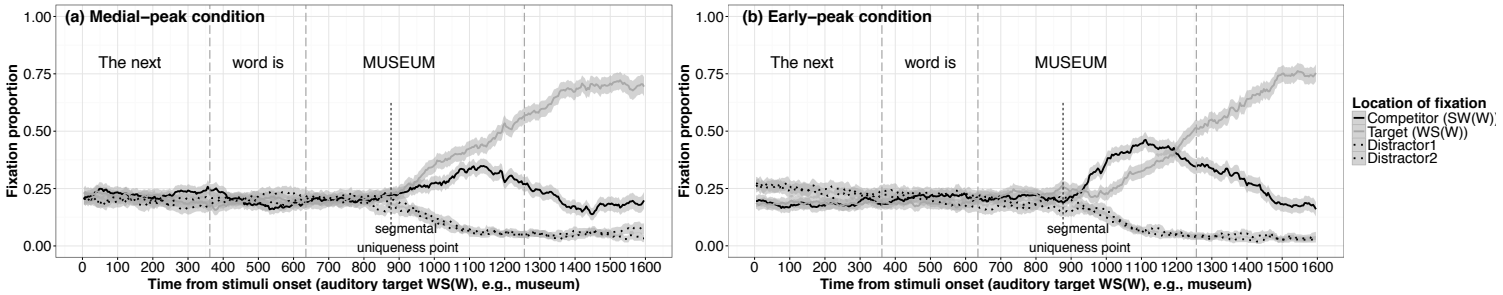
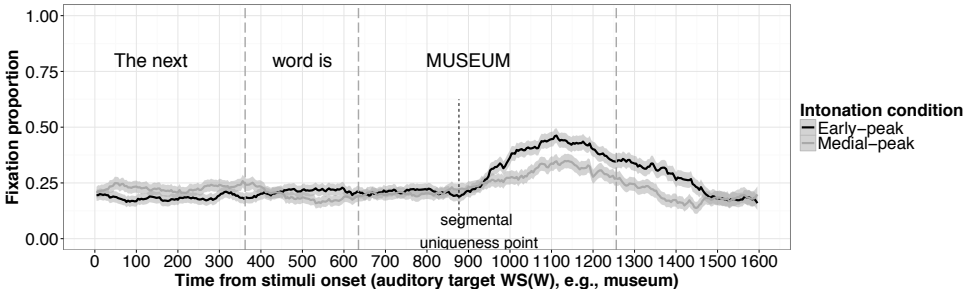


Figure 2. *Fixations to SW(W)-competitor (musical) in two intonation conditions; boundaries refer to acoustic landmarks.*



[1] Ladd, D. R. 1996. *Intonational phonology*. Cambridge: Cambridge University Press.

[2] Egger, S., & Braun, B. 2015. The role of pitch accents in the identification of metrical prominence. Poster at the *Int. Conference "Prominence in Language",* Cologne.

[3] Zahner, K., Schönhuber, M., Grijzenhout, J., & Braun, B. 2016. Pitch accent type affects lexical activation in German: Evidence from eye tracking. *Proceedings of the 16th Speech Science and Technology Conference,* Sydney, 185-188.

[4] Delattre, P. 1969. An acoustic and articulatory study of vowel reduction in four languages. *International Review of Applied Linguistics and Language Teaching (IRAL),* 7, 294-325.

[5] Cooper, N., Cutler, A., & Wales, R. 2002. Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners. *Language and Speech,* 45, 207-228.

[6] Cutler, A., & Pasveer, D. 2006. Explaining cross-linguistic differences in effects of lexical stress on spoken-word recognition. *Proceedings of the 3rd International Conference on Speech Prosody*, Dresden, 237-240.

[7] Braun, B., Lemhöfer, K., & Mani, N. 2011. Perceiving unstressed vowels in foreign-accented English. *Journal of the Acoustical Society of America,* 129, 376-387.

[8] Friedrich, C. K. 2002. *Prosody and spoken word recognition: Behavioral and ERP correlates.* PhD dissertation, University of Leipzig.

[9] Friedrich, C. K., Alter, K., & Kotz, S. A. 2001. An electrophysiological response to different pitch contours in words. *Neuroreport,* 12, 3189-3191.

[10] Cutler, A., Wales, R., Cooper, N., & Janssen, J. 2007. Dutch listeners' use of suprasegmental cues to English stress, in *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, 1913-1916.

[11] Reinisch, E., Jesse, A., & McQueen, J. M. 2010. Early use of phonetic information in spoken word recognition: Lexical stress drives eye movements immediately. *The Quarterly Journal of Experimental Psychology,* 63, 772-783.

[12] McGregor, J., & Palethorpe, S. 2008. High rising tunes in Australian English: The communicative function of L* and H* pitch accent onsets. *Australian Journal of Linguistics,* 28, 171-193.

Diachronically stable, lexically specific variation:
The phonological representation of secondary /æ/-lengthening

Thomas Kettig
University of Hawai'i at Mānoa

Secondary /æ/-lengthening, also known as the BAD-LAD split (*bad* as [bæːd] and *lad* as [læd]), is noted in several 20th century impressionistic descriptions of Southern Standard British English (SSBE) [4, 8, 12]. This paper reports results from an acoustic phonetic analysis of the duration of monosyllabic /æ/ (TRAP) words in SSBE. The apparent stability of this lengthening over the past century indicates that such sub-phonemic lexical variation is not necessarily indicative of a change still in progress. While its sub-phonemic nature is best accounted for with a usage-based phonological framework in which detailed phonetic properties are stored in lexical representations, the appearance of lengthening in higher-frequency words otherwise predicted to undergo reduction points to a more complex interaction of vowel duration and frequency than previously reported [7, 10].

Twenty-one native SSBE-speaking students at the University of Cambridge, aged 18–24, were recorded reading sentences embedded with 73 monosyllabic words containing the stressed vowel /æ/ (token n=1,790). Duration and F1/F2 were measured in Praat [1], with modal, breathy, and preaspirated sections coded separately and vowel duration defined as the modal plus breathy portions. A linear mixed effects model predicting duration with voicing and manner and place of articulation of coda consonants as well as word frequency entered as fixed effects along with by-word and by-speaker random intercepts and their interactions was run. Effects of phonological environment on length were observed as expected; results also indicated significant lengthening of higher-frequency words (p=0.016). The coefficients of the by-word random intercepts offer a window into how 'unexpectedly' long or short individual /æ/ words are once the phonological features of the post-tonic consonant(s) are accounted for; words singled out a century ago as being 'long' (e.g. *bag, that, bad, jam, sad*) still top the ranking (Figure 1). In a linear model, extreme coefficient scores – both high and low – also had a small but significant correlation with higher frequency (R^2=.090, p<0.001).

The secondary /æ/-lengthening reported here bears a striking resemblance to the reconstructed initial stages of primary /æ/-lengthening (also known as the TRAP-BATH split), a diachronic process that began in Southern England around the 17th century as allophonic length variation in reflexes of the Middle English /æ/ vowel [12]. While Lexical Phonology has been used to analyze primary /æ/-lengthening, such a model predicts that allophonic contrasts, processed post-lexically, should never be involved in lexically selective change [6, 9]. On the other hand, usage-based theories in which people store detailed phonetic properties of individual words allow persistent biases in pronunciation to accumulate into small but consistent sub-phonemic differences [2, 11]. However, while Exemplar Theory has been used to account for vowel *reduction* in higher-frequency words, the present data indicates that the opposing process of *lengthening* has in fact occurred some high-frequency words [3, 5]. It appears that just as primary /æ/-lengthening represents an ossified, half completed sound change, secondary /æ/-lengthening seems to have undergone the same lengthening process but lexically ossified at the sub-phonemic level. Phonemic contrastiveness does not seem to be a prerequisite for the stable maintenance of a lexically specified split, even when this directly counteracts an overarching articulatory process. Given this active resistance to the process of phonetic reduction, it is proposed that this in fact represents stable transmission and storage of lexically specified sub-phonemic allophony. In a usage-based framework, it would be predicted that such variation should affect more frequent rather than less frequent words, since a learner would have to encounter such words often enough to consistently store the duration differences in the word memories accessed in speech production.

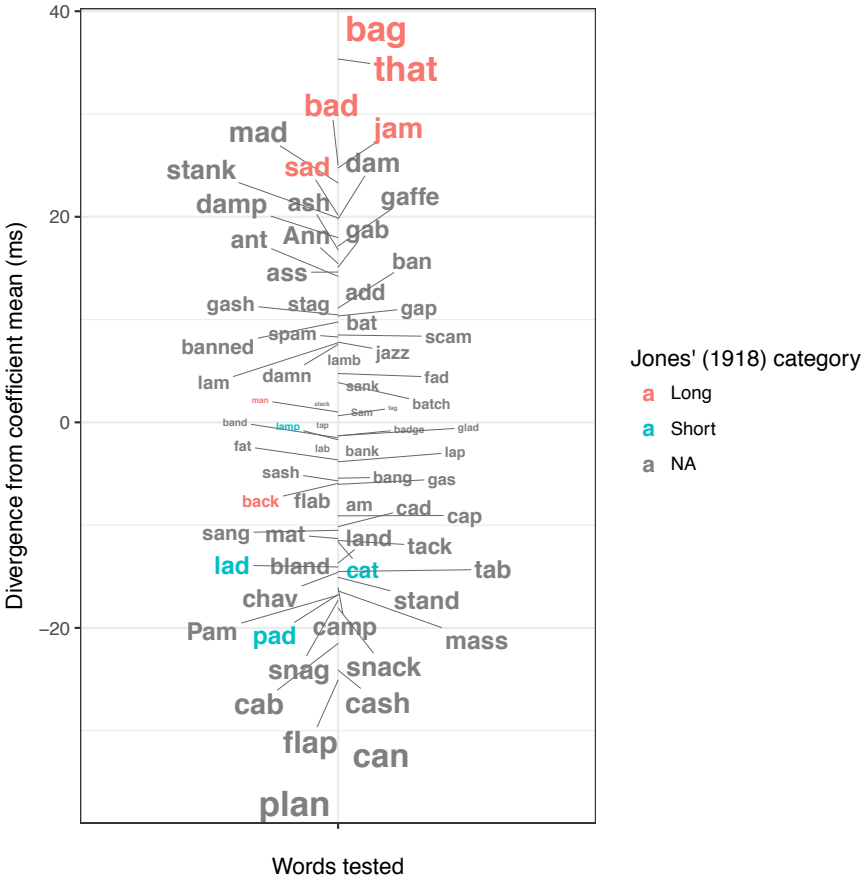


Figure 1. *Words investigated, plotted by amount of lengthening unaccounted for by fixed effects.*

[1] Boersma, P. & D. Weenink. 2015. *Praat: doing phonetics by computer* [Computer program]. Version 5.4.07. http://www.praat.org/.

[2] Bybee, J. 2000. The phonology of the lexicon: Evidence from lexical diffusion. In M. Barlow & S. Kemmer (Eds.), *Usage-based models of language*. Stanford: CSLI, pp. 65–85.

[3] Bybee, J. 2001. *Phonology and language use*. Cambridge: Cambridge University Press.

[4] Cruttenden, A. 2001. *Gimson's Pronunciation of English, Sixth Edition*. Oxford: Oxford University Press.

[5] Gahl, S. 2007. "Thyme" and "Time" are not homophones. Word durations in spontaneous speech. *Language* 84(3): 474–496.

[6] Harris, J. 1989. Towards a lexical analysis of sound change in progress. *Journal of Linguistics* 25(1): 35–56.

[7] Hay, J., J. Pierrehumbert, A. Walker, & P. LaShell. 2015. Tracking word frequency effects through 130 years of sound change. *Cognition* 139: 83–91.

[8] Jones, D. J. 1918. *An outline of English phonetics*. Leipzig: Teubner.

[9] Kiparsky, P. 1988. Phonological change, in F. J. Newmeyer (Ed.), *Linguistics: The Cambridge Survey. Volume I, Linguistic Theory: Foundations*. Cambridge: Cambridge University Press, pp. 363–415.

[10] Phillips, B. 2006. *Word frequency and lexical diffusion*. New York: Palgrave Macmillan.

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

[12] Wells, J. C. 1982. *Accents of English (Vols. 1-3)*. Cambridge: Cambridge University Press.

Monday, 13:15 - 14:15
Poster Session 1, #9

Non-local temporal adjustments caused by length contrasts: the case of Japanese

Takeki Kamiyama[1, 2] and Giuseppina Turco[2]
1. *LECSeL, EA 1569, Université Paris 8 Vincennes – Saint-Denis, France*
2. *Lab. de Phonétique et Phonologie (UMR 7018), CNRS / Sorbonne Nouvelle, France*

Non-local effects refer to a phenomenon by which non-adjacent segments within a certain domain (syllable, foot, word, etc.) may be similar with respect to some dimensions. Phonologically speaking, these effects may be also referred to as (vowel or consonant) *harmony*. While harmony phenomena often involve feature-based contrasts (e.g. the feature [nasal]) that reflect their basic defining articulatory properties, little is known about non-local effects induced by a contrast related to the temporal properties of the sound (see Turco & Braun [1] for a study on Italian). For Italian, Turco and Braun [1] found that the duration adjustment extended further to the word-initial consonant: the [p] in *palla* was significantly longer (about 9 ms) than that in *pala*. Sporadic data showing a similar tendency in Japanese has also been found in Han [2] and in Idemaru and Guion [3].

Similarly, this study deals with non-local duration differences caused by length contrast in another language containing a lexical contrast between geminate and singleton consonants: Japanese. Like in many world languages, Japanese has a lexical contrast between two types of consonants: singleton vs. geminate, as in the minimal pair *kako* “past” vs. *kakko* “parenthesis”. Phonologically, singleton and geminate consonants are distinguished by their association to one vs. two skeletal positions. Phonetically, the distinction between the two consonants is carried primarily by duration: geminates are systematically longer than their singleton counterparts (see, for instance, Ridouane [4]).

The goal of the study is to examine whether defining features associated with geminates are already anticipated on the non-adjacent word-initial consonant (e.g., whether in *kakko* the properties of the word-medial [kk] geminate are already foreshadowed on the word-initial [k] as compared to the [k] in the word *kako*). We seek to determine whether there are language-specific differences in the implementation of the onset consonant duration and on vowel-consonant timing coordination. Different from Italian, in Japanese the vowel before the upcoming geminate is longer than the vowel before the singleton consonant. It is hence possible that the vowel intervening between the influencing sound (the word-medial consonant, C2 henceforth) and the onset consonant (C1) may also play a role on the lengthening effect. As a matter of fact, the role of the intervening vowel is highly discussed in phonological accounts of harmony (cf. Odden [4]; Shaw [5]).

Five native speakers of Japanese (4 female and 1 male) from Tokyo and the surrounding areas read 20 minimal pairs (40 disyllabic words and nonse words: e.g. *kako* "past" vs. *kakko* "parenthesis") three times in a carrier sentence (/sore o __ toiu/) presented in Japanese orthography in a random order. The onset consonant of the target word (C1) was a stop (/p t k/) or a fricative (/s h/), and the word-medial one (C2) was a stop (/p t k b d ɡ/). Only words with non-high vowels /a e o/ were chosen since high vowels /i u/ are prone to undergo devoicing. The speakers were asked not to insert a pause before the target word. The duration of each consonant and vowel of the target word was measured in Praat (by the first author).

Preliminary results show that four of the five speakers produced significantly longer C1 when C2 was a geminate than when it was a singleton (respectively; $p < .05$, at most, in Student's t-test, see Figure 1), and one speaker showed a similar tendency (JP 4 in Figure 1).

In line with Turco and Braun [1], we believe that these duration differences may serve to anticipate and enhance the upcoming length contrast. However, other factors influencing the initial lengthening of the onset consonant, like voicing of the word-medial consonant, influence of the type of vowel (V2) and inter-speaker differences will be examined and discussed.

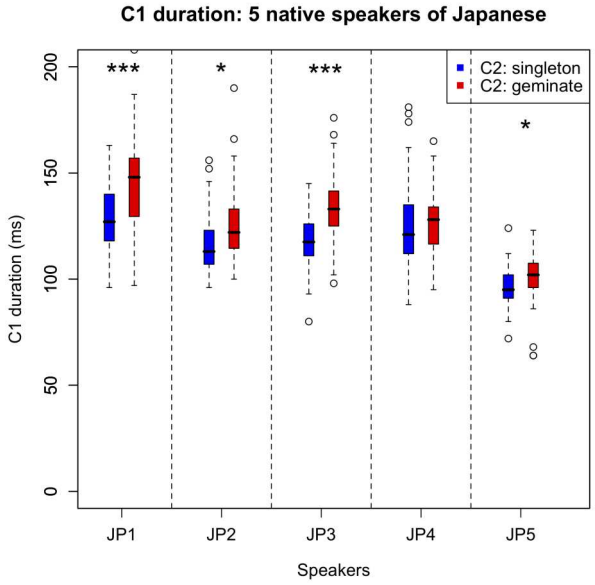


Figure 1. *(Mean) duration (in ms) of the initial consonant (C1) in Japanese singleton-geminate pairs (C2 stands for the word-medial consonant). Whiskers represent 2 standard deviations*.

[1] Turco, G., Braun, B. (2016). An acoustic study on non-local anticipatory effects of Italian length contrast. *Journal of the Acoustical Society of America*. 140(4), 2247-2256. doi:http://dx.doi.org/10.1121/1.4962982.

[2] Han, M. S. (1994). Acoustic manifestations of mora timing in Japanese. *Journal of the Acoustical Society of America* 96(1), 73-82.

[3] Idemaru, K. & Guion, S. G. (2008). Acoustic covariants of length contrast in Japanese stops. *Journal of the International Phonetic Association* 38(2), 167-186.

[4] Ridouane, R. (2010). Gemination at the junction of phonetics and phonology. In C. Fougeron, B. Kuehnert, M. D'Imperio & N. Vallée (Eds.), *Laboratory Phonology* 10, 61-90. Berlin: Mouton de Gruyter.

[5] Odden, D. (1994). Adjacency parameters in phonology. *Language* 70, 289-330.

[6] Shaw, P. A. (1991). Consonant harmony systems: The special status of coronal harmony. In C. Paradis & J.-F. Prunet (Eds.), *The Special Status of Coronals (Phonetics and Phonology 2)*, 125-157. San Diego: Academic Press.

Monday, 13:15 - 14:15
Poster Session 1, #10

Coordinative patterns underlying speech rhythm

Leonardo Lancia[1], Georgy Krasovitskiy[2], Franziska Stuntebeck[3]
[1]*Laboratoire de Phonétique et de Phonologie (CNRS, Sorbonne Nouvelle), France*,
[2]*University of Oxford, United Kingdom*, [3]*University of Zurich, Switzerland*.

The idea that languages can be sorted according to the rhythmic properties of the speech signal is supported by several perception studies conducted with stimuli deprived of their segmental content and their F0 modulations. It remains unclear however which properties of the acoustic signal allow speakers to make these distinctions [1]. One hypothesis is that rhythmic differences are due to differences in the coordination between the production of syllables and the production of prosodic prominence at the word level (e.g. [2], [3], [4]). This hypothesis is in line with Dauer's proposal [5] implying that, in languages traditionally considered as stress-timed (e.g. German or English), the production of word-level prominence imposes stronger constraints on the syllabic level than in languages traditionally considered as syllable-timed (e.g. French or Italian). However, available production studies either do not focus on typological dimensions or failed to report robust quantitative differences in line with perceptual data. In our study we analysed narratives produced by 30 speakers of five languages: English, German, French, Italian and Polish (ca. five speakers per language). Speakers were asked to describe the events occurring in the pear story video [6], which is often employed in comparing narrations across languages. In order to provide different conditions of enunciation, the participants were asked to tell the story twice: first online, during the projection of the video, and once more offline, after the projection. Our analyses followed [4] in that we estimated the coordination between the production of syllables and the production of prosodic prominence by measuring the coordination between the acoustic energy modulation due to the production of the syllables (***syllAM***) and the energy modulation due to the production of prominence (***stressAM***). These were obtained by band-pass filtering the amplitude envelope signal with cutoff frequencies at 2.3 Hz and 7 Hz for ***syllAM*** and at 0.8 Hz and 2.3 Hz for ***stressAM*** (see [4], for details). However, we characterise the level of coordination between these signals by adopting a Joint Recurrence Analysis approach [7], modified following [8] to deal with strongly non-stationary processes. Via this technique, we extract a coordination index (***CI***) that quantifies how often both the ***syllAM*** and ***stressAM*** simultaneously repeat behaviours that were jointly produced in the past. This value is normalized by the amount of repeated behaviour independently observed in each signal. From each recording, once removed pauses and hesitations, we obtained several ***CI*** values by analysing separately different portions of constant duration (16 ms). To exclude the hypothesis that observable cross-linguistic differences are due to the statistics of the languages under study, ***CI*** values obtained from the observed ***syllAM*** and ***stressAM*** were compared with those obtained from surrogate ***syllAM*** and ***stressAM*** constructed to display the same dynamics of the observed time-series without being temporally coordinated [9]. Mixed model regression (speaker and recording as random factors) shows that the level of coordination between syllables and prosodic prominence is significantly smaller in Roman languages than in German. This is true only for observed data and in both enunciation conditions. Post-hoc comparisons confirm that ***CI*** values are weaker in each Roman language compared to each Germanic language in both enunciation conditions but only in observed data. The ***CI*** values obtained from Polish differ only from the Italian values in the online condition and between conditions, and from the French values in the offline condition. In summary, the level of coordination between the production of syllables and the production of prosodic prominence at the word level appears to be a dimension pertinent to perceived cross-linguistic rhythmic differences. However, other dimensions may be required to account for the behaviour of Polish, whose rhythm is perceived as different also from that of German languages [10].

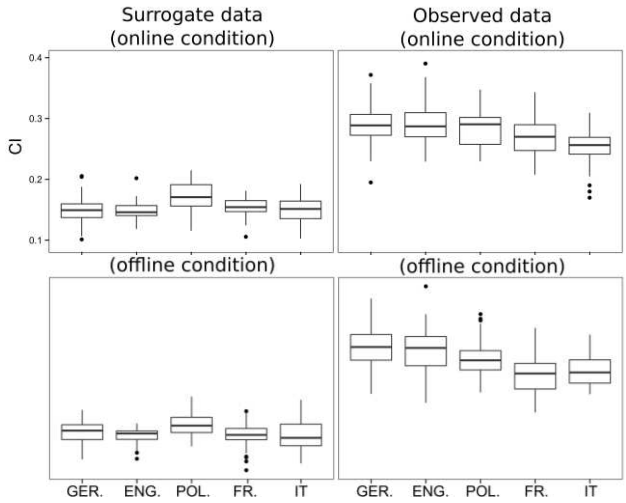


Figure 1. *CI values over languages grouped by data type and by enunciation condition. Left panels: surrogate data. Right panels: observed data. Top panels: online condition. Bottom panels: offline condition.*

References

[1] Arvaniti, A. (2012). The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics, 40*(3), 351-373.

[2] Barbosa, P. N. A. (2007). From syntax to acoustic duration: A dynamical model of speech rhythm production. *Speech Communication, 49*(9), 725-742.

[3] Goswami, U., & Leong, V. (2013). Speech rhythm and temporal structure: converging perspectives. *Laboratory Phonology, 4*(1), 67-92.

[4] Tilsen, S., & Arvaniti, A. (2013). Speech rhythm analysis with decomposition of the amplitude envelope: characterizing rhythmic patterns within and across languages. *The Journal of the Acoustical Society of America, 134*(1), 628-639.

[5] Dauer, R. M. (1987). Phonetic and phonological components of language rhythm. In *Proceedings of the 11th international congress of phonetic sciences*, Tallin, Estonia.

[6] Chafe, W. L. (1980). *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production*. Norwood, New Jersey: Ablex.

[7] Marwan, N., Romano, M. C., Thiel, M., & Kurths, J. (2007). Recurrence plots for the analysis of complex systems. *Physics Reports, 438*(5), 237-329.

[8] Lancia, L., Voigt, D., Krasovitskiy, G. (2016). Characterization of laryngealization as irregular vocal fold vibration and interaction with prosodic prominence. *Journal of Phonetics*, 54, 80-97.

[9] Thiel, M., Romano, M. C., Kurths, J., Rolfs, M., & Kliegl, R. (2006). Twin surrogates to test for complex synchronisation. *Europhysics Letters*, *75*(4), 535.

[10] Ramus, F., Dupoux, E., & Mehler, J. (2003). The psychological reality of rhythm classes: Perceptual studies. In *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona, Spain.

Monday, 13:15 - 14:15
Poster Session 1, #11

Light and dark /l/ in American English: The role of tongue narrowing

Matt Bauer
Illinois Institute of Technology

Among the variants of the lateral in American English, two have been most intensely studied: light /l/ at the beginning of syllables, as in “leap”, and dark /l/ at the end of syllables, as in “peel.” A number of researchers have argued that both light and dark /l/ have a primary coronal gesture and a secondary dorsal gesture, however, in light /l/, the coronal and dorsal gestures align, whereas in dark /l/, the dorsal gesture leads and the coronal gesture lags [1,2,3]. The prominence of the dorsal gesture in dark /l/ corresponds to its description of being velarized. These basic facts about gestural timing of light versus dark /l/ appear more or less consistent across dialects of English as well as in Catalan and German [4,5,6].

By one account [1], the reason for the difference in gestural timing for the primary and secondary articulation is due to the proximity of a gesture to nucleus or periphery of a syllable: dorsal gestures in liquids are open, and thus, vowel-like and align with the nucleus, while coronal gestures in liquids are constricting, and thus consonant-like and align with the periphery. Therefore, in /CV/ articulations, the coronal gesture of the /l/ leads, whereas in VC articulation, the dorsal gesture of the /l/ leads. An alternative explanation of the gestural differences between light and dark /l/ may have to do with resolving articulatory conflict. Laterals require the tongue body to narrow, to allow a channel of airflow around the margin(s) of the tongue [7]. Narrowing is achieved by drawing the margins of the tongue inward. However, any narrowing may conflict with keeping the mass of the tongue body in a position for vowel articulation. In such cases, articulatory conflict may be resolved with a loss of a narrowing gesture. For example, in CV sequences, where it is generally thought that gestures for the consonant and vowel are synchronous [8], a narrowing gesture is at odds with, and may yield to, the requirements of gestures for realizing vowels. In contrast, in VC sequences, where consonant gestures are separated from the onset of vowel gestures with an interval of time, a narrowing gesture would not conflict with vowel gestures. Because of the hydrostatic nature of the tongue, narrowing may automatically displace the dorsal volume of the tongue rearward, perhaps even leading to tongue tip retraction (i.e. vocalization) seen in some varieties of dark /l/.

At present, there is no description of the margins of the tongue during articulation of light or dark /l/--a crucial factor in the articulation of a lateral. The present study examined the margins of the tongue during the production of light and dark /l/ before and after /i/, /æ/, /u/, and /a/. Both midline and anterior views of the tongue were collected using ultrasound. Tongue narrowing was measured as the point in the lateral where the shortest distance obtained between the tongue margins (e.g., Figure 1). Results show that tongue narrowing for the lateral is less extreme in initial position compared to final position, consistent with the alternative view. Across all vowels, in post-vocalic position, the margins of the tongue are considerably more narrowed compared to pre-vocalic position. Moreover, the degree of tongue backing and tongue narrowing appear to be correlative. In other words, in CV articulations, where narrowing for the lateral is at odds with a forward tongue position for the vowel, particularly with /i/, the tongue does not narrow. In contrast, in VC sequences, where dorsal gestures for the vowel do not conflict with narrowing for the lateral, narrowing for the lateral is evident. The results may suggest the secondary gesture involves tongue narrowing (not just dorsum backing) in order to achieve a side channel required of laterals.

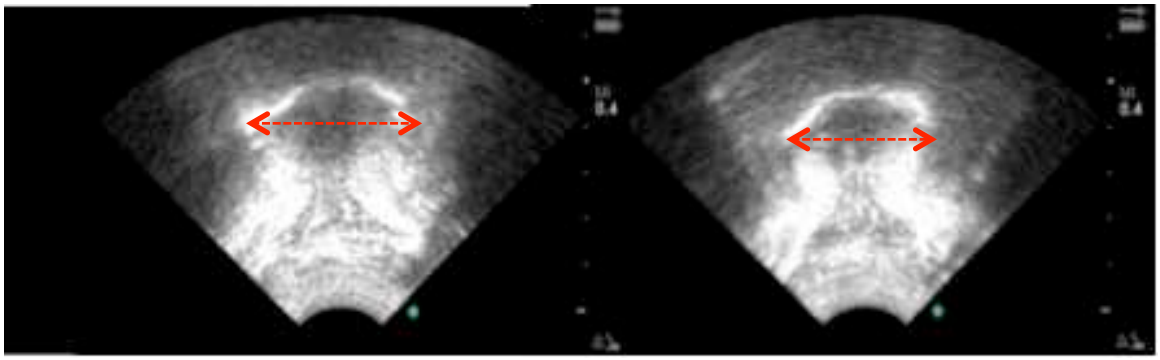


Figure 1. Anterior cross-section of tongue during the production of /l/ in “leap” (left) and “peel” (right). The distance, shown as dotted line, between the sides of the tongue is greater for the lateral in “leap” than it is for “peel.”

References

[1] Sproat, R., & Fujimura, O. (1993). Allophonic variation in English/l/and its implications for phonetic implementation. Journal of phonetics, 21(3), 291-311.

[2] Recasens, D. (2012). A cross-language acoustic study of initial and final allophones of/l. Speech Communication, 54(3), 368-383.

[3] Browman, C. P., & Goldstein, L. (1995). Gestural syllable position effects in American English. Producing speech: Contemporary issues, 19-33.

[4] Turton, D. (2015). Determining categoricity in English/l/-darkening: A principal component analysis of ultrasound spline data. Proceedings of the XVIIIth ICPhS.

[5] Recasens, D., Pallarès, M. D., & Fontdevila, J. (1998). An electropalatographic and acoustic study of temporal coarticulation for Catalan dark/l/and German clear/l. Phonetica, 55(1-2), 53- 79.

[6] Recasens, D., & Espinosa, A. (2005). Articulatory, positional and coarticulatory characteristics for clear/l/and dark/l: evidence from two Catalan dialects. Journal of the International Phonetic Association, 35(01), 1-25.

[7] Stevens, K. N. (2000). Acoustic phonetics (Vol. 30). MIT press.

[8] Browman, C. P., & Goldstein, L. (1990). Tiers in articulatory phonology, with some implications for casual speech. Papers in laboratory phonology I: Between the grammar and physics of speech, 341-376.

Lombard effect-related acoustic changes in the production of subsidiary stress

Beata Łukaszewicz, Ewa Zajbt, Urszula Krawczyk and Janina Mołczanow
University of Warsaw

The Lombard effect [1] is a well-known phenomenon whereby an increase in vocal amplitude occurs in response to the presence of background noise. Apart from intensity rise, vocalic adjustment taking place in Lombard speech includes heightened F0 values, a shift of spectral energy toward the medium frequencies, and longer duration. Most studies of Lombard speech thus far have concentrated on global speech reorganization occurring in noise conditions [2, 3, 4, 5, 6, 7, 8, 9]. However, a recent study of production of lexical stress in English [10] reveals that Lombard effect-related changes do not affect all syllables in the utterance equally: stressed syllables exhibit greater intensity rises than the immediately adjoining unstressed syllables. The increased contrast between stressed and unstressed syllables facilitates lexical identification in the presence of background noise. The aim of the current study is to test whether contrastivity enhancement during Lombard speech extends to grammatical stress systems, where the more conspicuous rhythmic structure can potentially improve overall intelligibility but has no effect on the meaning of words. If rhythmic stress is enhanced, the next question is whether this is effected in terms of the same acoustic parameters that are used to express rhythmic stress in the quiet condition.

We report on an acoustic study of Lombard-effect related changes in the production of subsidiary stress in Polish, which has main stress on the penult and secondary and tertiary stresses in word-initial and word-medial positions. The stimuli included 10 segmentally matched five- and six-syllable words (e.g. *ˌnagabyˈwany* 'chat up (PAST PART NOM SG)' – *ˌnagaˌbywaˈnego* 'chat up (PAST PART GEN SG)'). In the experiment, 20 participants produced the same words in a frame in three conditions: silence, multi-talker babble, and white noise. The noise conditions were played into closed headphones at the intensity level of 62 dB SPL. Three acoustic parameters were investigated: intensity, F0, and duration. Measurements were conducted in *Praat* [11]. In order to make our results comparable to those reported in [10], as well as to the previous ones on rhythmic stress in Polish [12], apart from raw measurements we also calculated vocalic and consonantal PVIs (Pairwise Variability Index; [13, 14, 15, 10]) for neighbouring stressed and unstressed syllables. The effects of noise and rhythmic stress on intensity, F0, and duration were analysed via a series of linear mixed-effects models fitted with random effects for Item and Speaker. The analyses were done in SPSS, version 23.

As regards the production of the English lexical stress during Lombard speech, increased F0 and intensity, but not elongated duration, were reported in [10]. Similarly, we found enhanced duration of syllables carrying subsidiary stress only in the silence condition, but not in the noise conditions. A significant decrease in contrastivity, found both for the white noise condition and the multi-talker babble condition, may be indicative of a general hindering effect of background noise upon temporal contrasts in speech. As to the remaining two acoustic cues, the analysis did not reveal a uniform pattern of contrastivity enhancement/reduction. For the multi-talker babble condition, there were no significant stress-related differences in relative F0 and intensity in comparison to the silence condition. In contrast, in the white noise condition there was a statistically significant rise in both PVI_ST and PVI_dB values observed across the second and third syllables in hexasyllabic words (the tertiary stress condition), compared to the silence and multi-talker babble conditions. This indicates a sharper F0 and intensity contrast decline in the context of white noise. Overall, it appears that Lombard speech boosts lexically relevant metrical features, while subtler rhythmic cues which guide speech processing in normal environment are lost in the presence of background noise.

[1] Lombard, E. 1911. Le signe de l'elevation de la voix [The sign of the rise of the voice]. *Maladies Oreille, Larynx, Nez, Pharynx*, 27, 101–119.
[2] Stanton, B. J., Jamieson, L. H., & Allen, G. D. 1988. Acoustic-phonetic analysis of loud and Lombard speech in simulated cockpit conditions. *Proceedings of ICASSP '88, the International Conference on Acoustics, Speech and Signal Processing*, 331–333.
[3] Van Summers, W., Pisoni, B., Bernacki, H., Pedlow, R., & Stokes, M. 1988. Effect of noise on speech production: Acoustic and perceptual analyses. *The Journal of the Acoustical Society of America*, 84, 917–928.
[4] Mokbel, C. 1992. *Reconnaissance de la parole dans le bruit: Bruitage/débruitage*. Unpublished doctoral dissertation. Ecole Nationale Supérieure des Télécommunications, Paris, France.
[5] Junqua, J-C. 1996. The Lombard reflex and its role on human listeners and automatic speech recognizers. *Speech Communication*, 20, 13-22.
[6] Castellanos, A., Benedi, J.M., & Casacuberta, F. 1996. An analysis of general acoustic-phonetic features for Spanish speech produced with the Lombard effect. *Speech Communication*, 20, 23-35.
[7] Kim, S. 2005. Durational characteristics of Korean Lombard speech. *Proceedings of Interspeech*, 2901–2904.
[8] Davis, C., Kim, J., Grauwinkel, K., & Mixdorff, H. 2006. Lombard speech: Auditory(A), Visual(V) and AV effects. *Proceedings of Speech Prosody*, 361–365.
[9] Garnier, M., Henrich, N., & Dubois, D. 2010. Influence of sound immersion and communicative interaction on the Lombard effect. *Journal of Speech, Language and Hearing Research*, 53, 588–608.
[10] Arciuli, J., Simpson, B. S., Vogel, A., & Ballard, K. J. 2014. Acoustic Changes in the Production of Lexical Stress during Lombard Speech. *Language and Speech*, 57(2), 149-162.
[11] Boersma, P., & Weenink, D. 1992–2015. *Praat: doing phonetics by computer* (version 6.0.08). Computer program.
[12] Łukaszewicz, B. 2015. Polish rhythmic stress revisited: phonetic evidence of an iterative system. Poster presented at *Phonetics and Phonology in Europe* (Cambridge, June 2015).
[13] Low, E. L., Grabe, E., & Nolan, F. 2000. Quantitative characterizations of speech rhythm: Syllable timing in Singapore English. *Language and Speech*, 43(4), 377–401.
[14] Ballard, K. J., Robin, D. A., McCabe, P., & McDonald, J. 2010. A treatment for dysprosody in childhood apraxia of speech. *Journal of Speech, Language, and Hearing Research*, 53, 1227–1245.
[15] Ballard, K. J., Djaja, D., Arciuli, J., James, D., & van Doorn, J. 2012. Developmental trajectory for production of prosody: Lexical stress contrastivity in children 3 to 7 years and adults. *Journal of Speech, Language, and Hearing Research*, 55, 1822–1835.

Monday, 13:15 - 14:15
Poster Session 1, #13

Production Planning and Frequency Effects in Sandhi: hiatus resolution in Spanish

Jeffrey Lamontagne and Francisco Torreira
McGill University

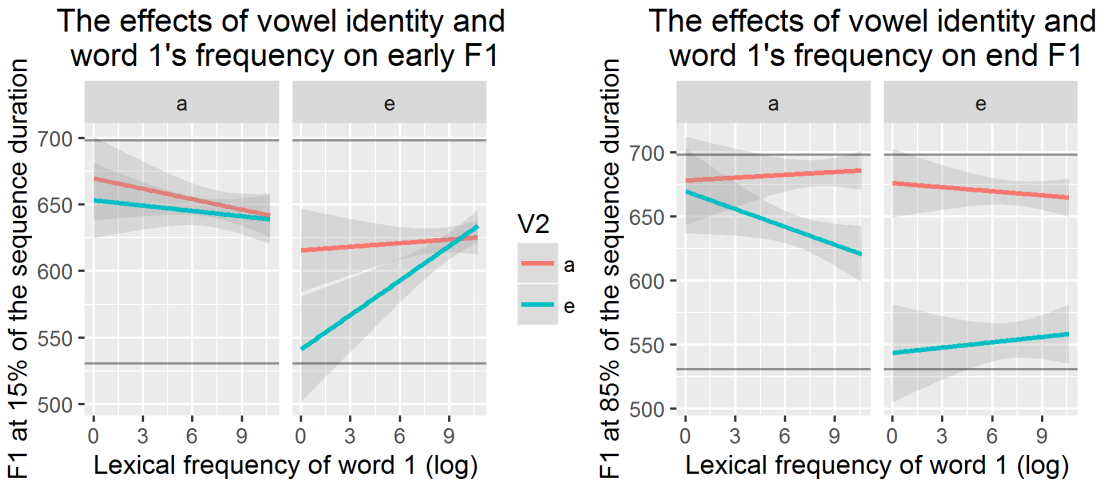
Production planning accounts of phonological variation predict that a cross-word alternation will occur more often when the words involved are easier to plan because that means that the speaker is more likely to know that the phonological context required for the alternation is present (e.g. Patterson and Connine 2001, Wagner 2012). For example, flapping a word-final /t/ before a vowel-initial word occurs more often when the second word is frequent because the speaker retrieves the following word earlier and has therefore planned the upcoming vowel early enough to articulate a flap instead of a stop (Wagner 2012). Previous analyses have focused on cases where a word's pronunciation is influenced by upcoming structure (e.g. Wagner 2012, Kilbourn-Ceron et al. 2016), like in the flapping example, without examining cases where variants could be sensitive to the previous word.

In this paper, we look at a case where the process could go in either direction: cross-word hiatus sequences in Spanish, as in *l**a e**scuela* 'the school', are sometimes realised as [a(ː)] (V2 deletes/assimilates following a vowel) and other times as [e(ː)] (V1 deletes/assimilates preceding a vowel), in addition to being faithfully realised as [ae] (e.g. Garrido 2013). There's a basic sonority effect in that /a/ tends to be deleted less than /e/, but either one can be deleted (Garrido 2013). Frequency effects are commonly observed in phonological phenomena, but there's typically no prediction that a word's pronunciation will be influenced by the frequency of another word (e.g. Mitterer and Russel 2013). We therefore test a prediction that's crucial to production planning: production planning predicts that word 2 being easier to plan (ex. more frequent) will be associated with more reduction in word 1.

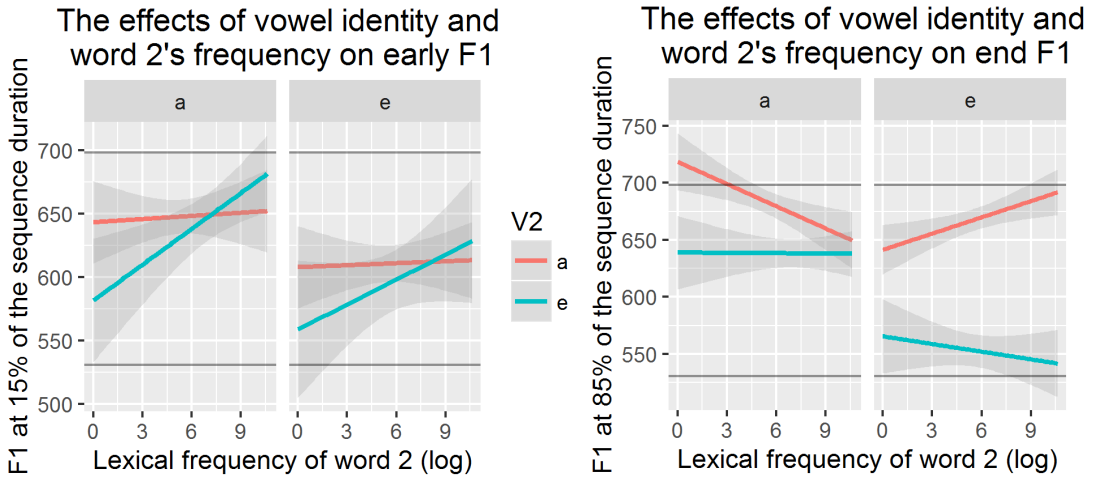
12 500 tokens of cross-word hiatus containing only /a/ and/or /e/ and without intervening punctuation in the orthography were extracted from the spontaneous speech of 35 native Spanish speakers in the Nijmegen Corpus of Conversational Spanish (Torreira and Ernestus 2010). The results were analysed using mixed-effect linear regression with by-word and by-speaker random slopes and intercepts to predict F1 at 15% and 85% of the vowel sequence's duration. A production planning account predicts that word 1's final vowel (V1) will be more reduced when word 1 is frequent (/ea/ > [a]) and that word 2's initial vowel (V2) will be more reduced when word 2 is frequent (/ea/ > [a]), as well as that word 1 and word 2 being more frequent both predict that V1 could be more reduced. We additionally examined the conditional probability of word 2 given word 1 as more likely combinations are associated with faster lexical access and increased phonological planning (Kilbourn-Ceron et al. 2016).

Our results reveal that at the start of the vowel sequence both V1 ($p=0.003$) and V2 ($p=0.002$) significantly affect F1. Crucially, we find that there's a significant – but smaller – interaction between the second vowel and the first word's frequency that increases the effect of that upcoming vowel ($p=0.024$). Late into the vowel sequence, we see that again both V1 ($p<0.001$) and V2 ($p=0.048$) are significant, but that the interaction between the vowel identities now overtakes the effect of either vowel in size ($p<0.0001$). There's an additional significant effect of the sequence duration ($p<0.0001$) and interaction between V1 and word 1's frequency ($p<0.0001$). Word 2's conditional probability was additionally a significant predictor of early F1, consistent with the predicted effect of easier lexical access.

Overall, our results are consistent with a production planning account given how the word frequencies and conditional probabilities influence the likelihood of resolving vowel hiatus. We additionally find evidence of more [a]-like pronunciations being favoured, consistent with Garrido's (2013) earlier observations. Finally, we show asymmetries between the left-to-right and right-to-left applications of hiatus resolution, consistent with production planning predictions.



Figures 1 (left) and 2 (right): *The effects of vowels' stress and word 1's lexical frequency of the first formant at 15% (left) and 85% (right) of the vowel's duration. The top grey line identifies /a/'s mean F1, while the bottom one identifies /e/'s mean F1.*



Figures 3 (left) and 4 (right): *The effects of vowels' stress and word 2's lexical frequency of the first formant at 15% (left) and 85% (right) of the vowel's duration. The top grey line identifies /a/'s mean F1, while the bottom one identifies /e/'s mean F1.*

References

Garrido, M. 2013. "Hiatus Resolution in Spanish: Motivating Forces, Constraining Factors, and Research Methods". *Language and Linguistics Compass* 7: 339–350.

Kilbourn-Ceron, O., M. Wagner, M. Clayards. 2016. "The effect of production planning locality on external sandhi: a study in /t/". *Proceedings of the 52nd Meeting of the Chicago Linguistics Society*.

Mitterer, H. and K. Russel. 2013. "How phonological reductions sometimes help the listener". *Journal of Experimental Psychology* 39 (3): 977-984.

Patterson, D., and C. M. Connine. 2001. "Variant frequency in flap production: A corpus analysis of variant frequency in American English flap production". *Phonetica* 58:254-275.

Torreira, F. and Ernestus, M. 2010. "The Nijmegen Corpus of Casual Spanish". *Proceedings of LREC 2010*.

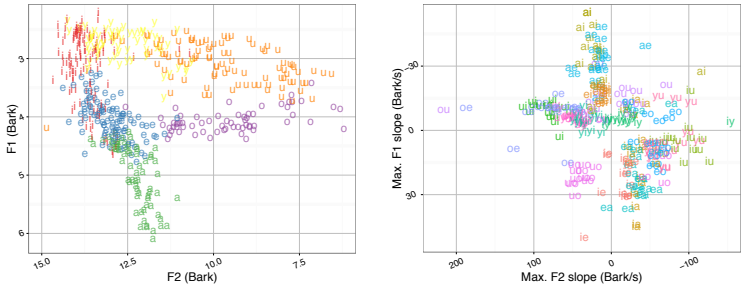
Wagner, M. 2012. "Locality in phonology and production planning". *McGill Working Papers in Linguistics* 23.1.

A comparison of the information conveyed by static and dynamic formant cues for vowel classification in vowel-vowel sequences.

Crouzet, Olivier & Millasseau, Julien
Laboratoire de Linguistique de Nantes (LLING), UMR6310 – Université de Nantes / CNRS, Chemin de la Censive du Tertre, 44312 Nantes Cedex

Framework: Various approaches to vowel category specification have been proposed, among which proposals concerning the respective contributions of static and dynamic cues were conceived that date back to Strange et al.'s seminal review [1]. These views are still very influent nowadays and most contemporary results have evidenced the presence of both static and dynamic cues to the classification of vowel categories [2, 3, 4]. Dynamic cues in this respect have been defined as "vowel inherent spectral changes"(VISC), underlying the understanding that they should characterize a single segment category. Carré [5] proposed an alternative view according to which *formant transition fastest slopes* measured in the time interval during which the vocal tract goes from one configuration to the other would constitute better candidates to reflect a consistent distinction between phonological vowel categories ([5]; see also [6], for related arguments based on perceptual data). Though this conception may seem relatively in line with previous dynamical views of vowel specification, it has strong consequences on the understanding of the relation between articulatory / acoustic phenomena and the phonological representations that would be involved in speakers / listeners. Indeed, this approach predicts that "single-segment" vowel categories or related combinations of primitives (e.g. VISC, [3]) may not be an appropriate level of representation. Rather, transitions "between" segments (therefore *diphones*) would constitute a more adequate level of categorization and phonological organization.

Data: Carré's arguments [5] were based on unidimensional measurements only. Also, speech rates were not controlled formally. Therefore, an experiment was conducted in which French V_1V_2 vowel sequences were collected in short "determiner–noun–adjective" sequences at three different target speech rates using a dedicated computer interface. A statistical classification process (Linear Discrimant Analysis) was then performed on both static (middle formant frequencies) and dynamic (maximal formant transition slopes) measurements using various combinations of F_1, F_2, F_3 and $f0$. Overall, 1440 V_1V_2 utterances (2880 single vowels) were collected from 5 speakers and were passed through LDA modelling. Specific data treatments (z-score normalisation and spline smoothing) were applied to satisfy requirements for the extraction and modelling of both frequencies and maximal slopes. Classification performance was evaluated as a function of the number of available categories [7]. At first glance, a classical $F_1 \sim F_2$–space representation clearly favors the static hypothesis (Fig. 1). Similarly, comparing coefficients of variation for individual parameters results in much higher dispersion for formant velocity than for static measures. However, multidimensional classification analyses provide more measured views of the articulation between velocity and classification, showing that according to classification performance, velocity cues compete favorably with static cues (Tab. 1–2). Further analyses are being conducted in order to estimate measurement variation. These results will be discussed in light of current approaches to dynamical systems theory in phonetics and phonology [8, 9].



(a) (b)

Figure 1: Graphical representation (for Speaker 2 only, due to space limitations) of the $F1 \sim F2$ plane (a) of static formant frequencies taken at 50% of the total duration of the vowel (V_1 and V_2, in Bark) ; (b) of maximum formant velocity (from V_1 to V_2, in Bark/s).

Table 1: Measurements of LDA classification performance based on static formant frequencies ($[\frac{p-p_{ref}}{1-p_{ref}} \times 100$, with $p_{ref} = 1/6$, [7]).

	/i/	/e/	/a/	/o/	/u/	/y/	**global**
$F1+F2$	58.0	48.2	53.3	46.2	56.5	2.6	**44.1**
$F1+F2+F3$	51.9	47.2	53.6	46.2	53.3	40.9	**48.9**
$F1+F2+F3+f_0$	51.9	49.1	53.9	46.2	52.9	40.9	**49.2**

Table 2: Measurements of LDA classification performance based on dynamic (maximal slope) formant frequencies ($\frac{p-p_{ref}}{1-p_{ref}} \times 100$, with $p_{ref} = 1/16$, [7]).

	/ie/	/ei/	/ia/	/ai/	/iu/	/ui/	/iy/	/yi/	
$F1+F2$	9.3	70.0	-1.6	57.7	65.6	39.3	16.4	40.8	
$F1+F2+F3$	20.0	70.0	13.6	68.0	57.2	48.5	53.3	59.6	
$F1+F2+F3+f_0$	38.7	70.0	38.9	66.9	53.6	55.4	54.9	59.6	
	/ea/	/ae/	/eo/	/oe/	/ou/	/uo/	/uy/	/yu/	**global**
$F1+F2$	60.9	24.3	42.0	17.1	-1.1	49.2	17.7	31.8	**33.7**
$F1+F2+F3$	63.5	41.0	39.4	37.4	10.0	41.4	44.7	43.3	**44.4**
$F1+F2+F3+f_0$	66.1	42.1	36.8	33.8	5.8	38.8	46.0	34.4	**46.4**

References

[1] Winifred Strange, JJ Jenkins, and TL Johnson. Dynamic specification of coarticulated vowels. *The Journal of the Acoustical Society of America*, 74(3):695–705.

[2] C. I. Watson and J. Harrington. Acoustic evidence for dynamic formant trajectories in australian english vowels. *The Journal of the Acoustical Society of America*, 106:458–468, 1999.

[3] G. S. Morrison. Theories of vowel inherent spectral change. In G. S. Morrison and P. F. Assman, editors, *Vowel Inherent Spectral Change*, pages 31–47. Sringer, Berlin, DE, 2013.

[4] Jaydene Elvin, Daniel Williams, and Paola Escudero. Dynamic acoustic properties of monophthongs and diphthongs in western sydney australian english. *The Journal of the Acoustical Society of America*, 140(1), 2016.

[5] René Carré. Signal dynamics in the production and perception of vowels. In F. Pellegrino, E. Marsico, I. Chitoran, and C. Coupé, editors, *Approaches to Phonological Complexity*, pages 59–81. Mouton de Gruyter, Berlin – New-York, 2009.

[6] Pierre Divenyi. Perception of complete and incomplete formant transitions in vowels. *Journal of the Acoustical Society of America*, (3):1427–1439.

[7] J. L. Schwartz, F. Berthommier, and C. Savariaux. Seeing to hear better: evidence for early audio-visual interactions in speech identification. *Cognition*, 93:B69–B78, 2004.

[8] Adamantios I Gafos and Stefan Benus. Dynamics of phonological cognition. *Cognitive science*, 30:905–943, 2006.

[9] B Tuller, P Case, M Ding, and J a Kelso. The nonlinear dynamics of speech categorization. *Journal of experimental psychology. Human perception and performance*, 20(1):3–16, 1994.

Multimodal prominences: Exploring the interplay and usage of focal pitch accents, eyebrow beats and head beats in Swedish news readings

Gilbert Ambrazaitis[1] and David House[2]
[1]*Lund University, Sweden,* [2]*KTH Stockholm, Sweden*

Facial beat gestures align with pitch accents in speech [1,2], and there is evidence suggesting that they are more likely to occur with *perceptually strong* accents than with *weak* ones [3]. The results of [3] also suggest equivalent, cumulative prominence functions of head and eyebrow movements. Furthermore, perception experiments [4] have shown how the two modalities (gestures, pitch accents) *interact* in the coding of *contrastive focus (vs. information focus),* where head beats were found to be more informative than eyebrow movements for the identification of contrast. However, we still lack essential knowledge as to whether and how head and eyebrow beats might combine with pitch accents in different ways to construct *different types of multimodal prominence (MMP),* and whether different MMPs might be used differentially in spoken communication. Do they, for instance, help to encode further *information-structural* dimensions such as the *theme/rheme* distinction [5]?

The present study attempts to approach this research question by (1) investigating how MMPs are constructed by combining eyebrow (EB) and head (HE) beats with so-called *focal accents* (FA, i.e., *sentence-level* pitch accents) in Stockholm Swedish (H*LH- for Accent 2, (H+)L*H- for Accent 1, where the final high tone H- characterizes the accent as sentence-level [6,7]), and (2) exploring patterns of usage of the (combined) prominence markers.

Our corpus so far consists of 31 brief news readings from Swedish Television, comprising speech from four news anchors (two female, two male) and 986 words (6 ½ minutes) in total. It was annotated for FA, HE beats and EB beats, independently by three annotators (Fleiss' Kappa (κ) for FA: $\kappa = 0.77$; HE: $\kappa = 0.69$; EB: $\kappa = 0.72$). A word was annotated for a (HE or EB) beat in the event that the head or at least one eyebrow rapidly changed its position, roughly within the temporal domain of the word. Beats were annotated for the entire corpus, irrespective of FA annotations, and vice versa.

Four types of (combined) prominence markers occur rather frequently in the corpus: FA+HE+EB (39 tokens), FA+HE (126), FA only (i.e., no gesture: 128), and HE only (58), while FA+EB (3) is pretty rare, and both HE+EB (10) and EB (15) in 70-80% of the cases occur on words directly preceding words annotated for FA(+HE), seemingly as a kind of upbeat (cf. the alignment measurements by [2]). That is, EB beats tend to occur primarily in combination with both FA and HE, i.e. seldom alone or with either HE or FA only.

A functional analysis of the data has revealed the following general tendencies: (1) In almost all noun phrases (and many verbs) at least one word receives some kind of prominence. A high density of prominent words is most probably a general characteristic trait of news speech, where almost all referents represent new information. (2) In sections representing the *rheme* of the news story, FA, HE, and FA+HE are all common prominence markers. (3) However, initial clauses, when presenting a *common ground* [8,9] or the *theme* (about 2/3 of the texts), we observe a preference for using FA only for the first 2-3 prominent words of the news story (see Fig. 1). (4) Finally, EB seems to be used mainly as a kind of intensification marker and is able to cancel out the observation in (3) when occurring early in a text (e.g. at 7 a.m. in Fig. 1). Typically, FA+HE+EB occurs on words that are especially noteworthy, often representing an overt contrast (e.g., 'despite'), but also magnitude, a high value (e.g., 'jewels worth about three hundred million kronor'), or emotionally loaded words (e.g., 'luckily'). These results on EB beats might be slightly at odds with the perceptual finding cited above [4], but in line with [2] discussing the use of EB raising for emphasis. Our results provide some initial evidence for a differential usage of the two visual prominence markers, rather than an equivalent and cumulative one, cf. our discussion of [3].

Time	***Stroke***	***är***	***den***	***tredje***	***vanligaste***	***dödsorsaken***	***i***	***Sverige.***
(a.m.)	'Stroke	is	the	third	leading	cause of death	in	Sweden.'
6.30	FA			FA				
7.00	FA HE EB			FA HE	FA	FA		FA
7.30	FA				FA	FA		
8.30	FA			FA				
					HE	HE		
9.00	FA			FA		FA		
					HE	HE		

Figure 1. *An example of the data: Annotations of focal pitch accents (FA), head beats (HE) and eyebrow beats (EB), aligned with corresponding words, for five repetitions of the initial sentence of the same news story from different times of the day, read by the same male news anchor; the sentence establishes a common ground for the remainder of the news story. Broadcast from May 8, 2013, Swedish Television.*

[1] Yasinnik, Y., Renwick, M., & Shattuck-Hufnagel, S. 2004. The timing of speech-accompanied gestures with respect to prosody. *Proceedings of From Sound to Sense* (MIT, Cambridge, MA), 97-102.

[2] Flecha-García, M. L. 2010. Eyebrow raises in dialogue and their relation to discourse structure, utterance function and pitch accents in English. *Speech communication,* 52, 542-554.

[3] Swerts, M., & Krahmer, E. 2010. Visual prosody of newsreaders: Effects of information structure, emotional content and intended audience on facial expressions. *Journal of Phonetics,* 38, 197-206.

[4] Prieto, P., Puglesi, C., Borràs-Comes, J., Arroyo, E., & Blat, J. 2015. Exploring the contribution of prosody and gesture to the perception of focus using an animated agent. *Journal of Phonetics,* 49(1), 41-54.

[5] Jackendoff, R.S. 1972. *Semantic interpretation in generative grammar*. Cambridge, Mass.: M.I.T. Press.

[6] Bruce, G. 1977. *Swedish Word Accents in Sentence Perspective*. Travaux de l'institut de linguistique de Lund 12. Malmö: Gleerup.

[7] Bruce, G. 2007. Components of a prosodic typology of Swedish intonation. In T. Riad & C. Gussenhoven (Eds.), *Tones and Tunes – Volume 1: Typological Studies in Word and Sentence Prosody*, pp. 113–146. Berlin; New York: Mouton de Gruyter.

[8] Clark, H. H., & Carlson, T. B. 1981. Context for comprehension. In J. Long & A. Baddeley (Eds.), *Attention and performance IX*. Hillsdale, NJ: Erlbaum.

[9] Clark, H. H., and Marshall, C. R. 1981. Definite reference and mutual knowledge. In A. K. Joshe, B. Webber, & I. A. Sag (Eds.), *Elements of discourse understanding*. Cambridge: Cambridge University Press.

Dialectal variation in prosodic timing and prosody-segment interactions

Rachel Smith[1] & Tamara Rathcke[2]
[1]*University of Glasgow,* [2]*University of Kent*

Duration is a phonetic marker of various segmental contrasts, and at the same time, signals prosodic boundaries and prosodic prominence. Given extensive cross-dialectal variation in both prosody and segments, it is interesting to explore how prosodic and segmental timing and their interactions vary across dialects. In this paper, we compare durational patterning in Standard Scottish English (SSE) and Standard Southern British English (SSBE). These dialects differ in their prosodic systems, and also in their segmental systems, especially their stressed vowels [1]. Duration is heavily implicated in the segmental variation: SSBE tense and lax vowel pairs contrast in both duration and quality, while the tense-lax opposition has no clear application in SSE; and SSE vowel durations are affected by the Scottish Vowel Length Rule (e.g. [8]), while SSBE vowel durations are sensitive to a following consonant's voicing. Corpus studies indicate that durations may be affected in complex ways by a word's number of syllables and its position in prosodic structure [6, 9], but these factors have not been systematically explored with controlled data.

We investigated segmental duration in trochaic disyllables (e.g. *cheesy, picky, hazy*) in speakers of SSE from Glasgow and of SSBE from Cambridge (henceforth, "Glasgow" and "SSBE"). We manipulated three factors which we expected to cause vowels to lengthen (predictions in parentheses): (1) vowel quality (tense > lax vowels, for SSBE only; low > mid > high vowels for both dialects); (2) phonological context (vowels longer before a voiced fricative or /r/, vs. a voiceless stop); (3) prosodic condition (vowels longer in prominent vs. non-prominent positions; vowels longer in phrase-final vs. -medial position)

The tokens were recorded in three prosodic conditions, e.g.:

- *I reckoned it was kind of* ***cheesy*** (nuclear final, *nf*);
- *I don't think it's* ***remotely*** *cheesy* (post-nuclear final, *pf*);
- *That sounds a bit too* ***cheesy*** *for them* (nuclear non-final, *nn*).

Target words were embedded in short scripted dialogues, read by the participant and a trained speaker of their dialect. We present data from five speakers of each dialect (n=4015 tokens).

Prosodic timing patterns were found to be dialect-dependent (Figure 1). Phrase-final lengthening (FL) was measured by comparing nuclear final with nuclear non-final words, and accentual lengthening (AL) was measured by comparing nuclear final with post-nuclear final words. FL occurred in both dialects, starting from the onset of the target word's stressed syllable, and increasing in magnitude as segments approached the phrase boundary. Intriguingly, however, AL differed dramatically across the dialects. In SSBE we found robust evidence of AL on all segments in the word, whereas in Glasgow AL only affected the initial consonant, and did not occur on the target word's stressed vowel or subsequent segments. We explore how these findings may relate to the different intonation contours typical of the two dialects.

We also observed prosody-segment interactions. For SSBE, significant accentual lengthening was found only for tense, and not for lax vowels, perhaps because the latter are constrained segmentally to be short. The tense-lax distinction does not apply in Glasgow, so this comparison cannot meaningfully be made, but ongoing work explores whether height constrains prosodic lengthening in Glasgow in a similar manner to tense-lax in SSBE. We model vowel durations as a function of their F1 values in order to explore whether high (intrinsically shorter) vowels are less susceptible to prosodic lengthening than low (intrinsically longer) vowels.

The data confirm that prosodic timing can differ dramatically across dialects, though our predictions in (1)-(3) do not find a straightforward support. Remarkably, Glaswegian English appears to belong to a small group of languages that show no or very little accentual lengthening (e.g. [3, 7]) whereas SSBE patterns align with the typical lengthening alternations found in other varieties of English and other languages. We explore several possible explanations for the results, including in terms of coupling relations among prosodic gestures, as proposed in the framework of Articulatory Phonology [2, 5]. We discuss how the data may inform the development of multi-dimensional models of speech timing.

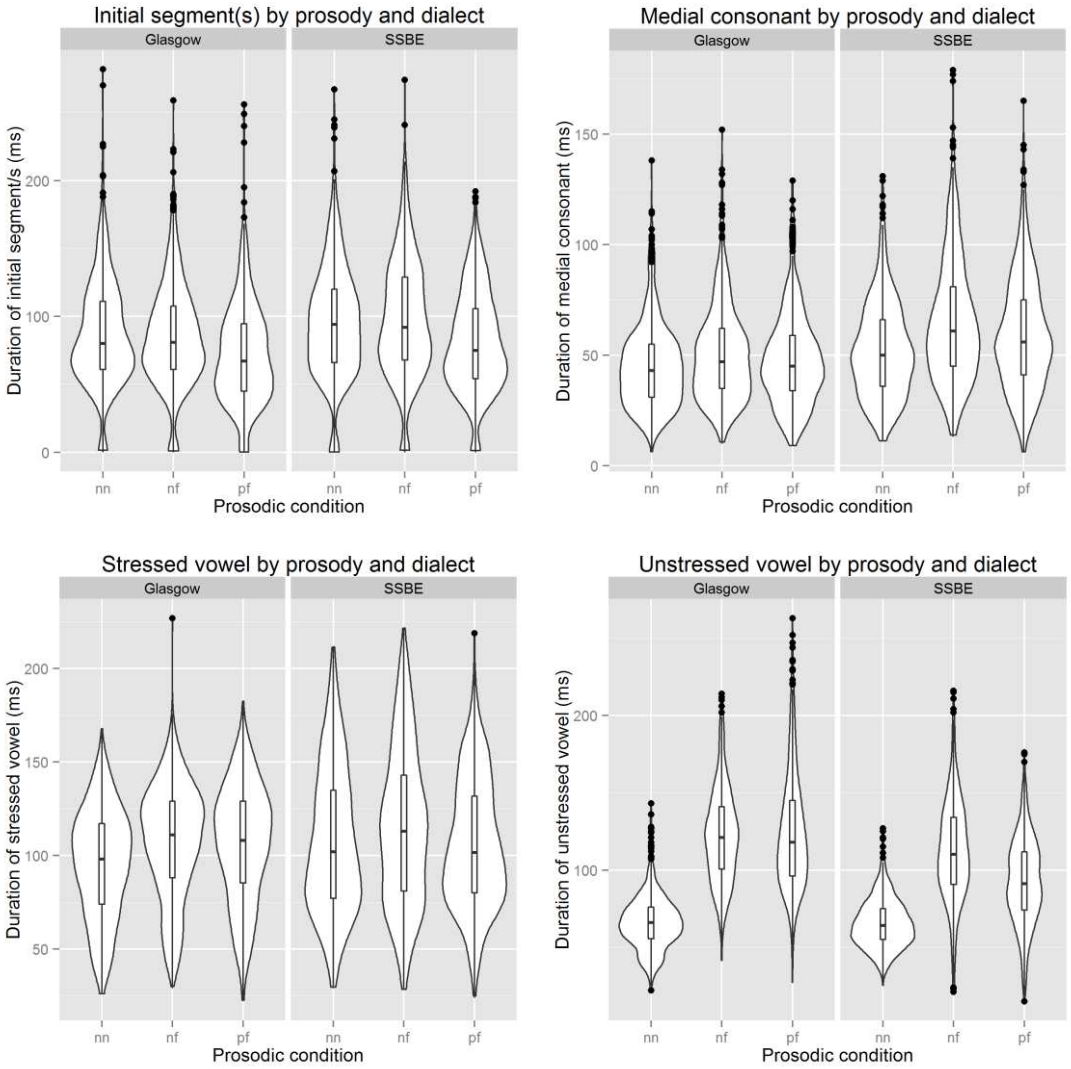


Figure 1. Segment durations in trochaic disyllables, according to prosodic condition (*nn, nf, pf*).

References

[1] Abercrombie, D. (1979). The accents of standard English in Scotland. In A. J. Aitken &. T. McArthur. (Eds.), *Languages of Scotland* (pp. 68-84). Edinburgh: Chambers.

[2] Byrd, D., & Riggs, D. (2008). Locality interactions with prominence in determining the scope of phrasal lengthening. *Journal of the International Phonetic Association*, 38, 187-202.

[3] Fletcher, J., & Evans, N. (2002). An acoustic intonational study of intonational prominence in two Australian languages. Journal of the International Phonetic Association 32, 123-40.

[4] Harrington, J., Fletcher, J. & Beckman, M. (2000) Manner and place conflicts in the articulation of accent in Australian English. In Broe & Pierrehumbert (eds.), *Papers in Laboratory Phonology V*, 40–51.

[5] Katsika, A. (2012), *Coordination of prosodic gestures at boundaries in Greek*, PhD thesis, Yale Univ.

[6] Rathcke, T. & Stuart-Smith, J. (2015). On the tail of the Scottish Vowel Length Rule. *Language and Speech* 59.

[7] Ortega-Llebaria, M., & Prieto, P. (2007). Disentangling stress from accent in Spanish: Production patterns of the stress contrast in deaccented syllables. In P. Prieto, J. Mascaró and M.-J. Solé (Eds.), *Segmental and Prosodic issues in Romance Phonology*. Amsterdam: John Benjamins, pp. 155-176.

[8] Scobbie, J. M., Hewlett, N., & Turk, A. (1999). Standard English in Edinburgh and Glasgow: the Scottish Vowel Length Rule revisited. In Foulkes, P. & Docherty, G. (eds.), *Urban voices: Accent studies in the British Isles* (pp. 230-245). London: Hodder.

[9] Scobbie, J.M., & Stuart-Smith, J. (2008). Quasi-phonemic contrast and the fuzzy inventory: examples from Scottish English. In Avery, P., Dresher, B.E., & Rice, K. (eds.), *Contrast in phonology: Theory, perception, acquisition*. Berlin: Mouton de Gruyter, 87-113.

Diachronic change and synchronic variation in Dutch vowel-/l/ sequences: the role of phonetics, phonology, and sociolinguistics

Cesko Voeten
Leiden University

The vowel system of Dutch is both synchronically and diachronically variable. Diachronically, two sound changes have taken place over the past decades: the diphthongization of [eː,øː,oː] towards [ei,øy,ou] ([1], [2], [3]) and the lowering of the original diphthongs [ɛi,oey,ɔu] towards [ai,ɒy,ɑu] ([4], [5]). Synchronically, these changes have led to regional differences ([6], [7], [8]). This latter point will be the focus of the talk: what is the regional variation and spread in the realizations of the tense mid vowels and diphthongs? The specific phenomena investigated in this study are (1) vowel diphthongization in front of non-approximant consonants; (2) diphthong neutralization in front of coda /l/; (3) retraction (of the tongue body) in front of coda /l/. These define the on-going diachronic changes in Dutch.

The above question is investigated using a large dataset called the 'teacher corpus' ([9]). The data were recorded (by [9]) via sociolinguistic interview in eight different regions in the Netherlands and Flanders. The regions were carefully selected so that for both countries the sampled regions consist of one central region (NL: the Randstad, FL: Brabant), one intermediate region (NL: South-Gelderland; FL: East-Flanders), and two peripheral regions (NL: Groningen and Dutch Limburg; FL: Flemish Limburg and West-Flanders). The corpus is further subcategorized for sex and age, and then has five speakers per cell, making for a total of 160 speakers. The subset of the dataset relevant to the study comprises a total of 5,407 tokens of 21 monosyllabic words (in total), in which 1,732 vowel tokens appear in front of nonapproximant consonants and 3,331 vowel tokens appear in front of coda /l/.

The available sound files had been fully segmented according to the manual transcription guidelines of the IFA protocol ([10]). While the boundaries between vowels and nonapproximant consonants were in all cases easy to identify, the segmentation between a vowel and a coda /l/ was in many cases difficult to verify, especially if the /l/ was heavily vocalized (cf. [11]). For this reason, it was decided to impute these vowel-consonant boundaries, by substituting the average duration of the other (reliably-segmented) vowels. To prevent this procedure from introducing undue bias into the results, the same procedure was then also applied to all other vowels, resulting in a consistent measurement interval starting from the onset of the vowel for every token.

The tense mid vowels and diphthongs were extracted from the sound files using Praat ([12]). For ease of statistical analysis, the method adopted extracted F1 and F2 values at ¾ of the measurement interval, and subtracted from this measurement the same measure at ¼ of the vowel, which resulted in a 'delta formant' measure. This greatly simplified the statistical analysis, and mitigated any effect of durational differences that was lost from the data by the imputing procedure described in the previous paragraph. The F1 and F2 delta formants were analyzed by means of a mixed-effects regression model (random slopes for subjects and items in a diagonal covariance configuration, cf. [13], [14]).

Results for the F1 show significant regional variation in the realizations of the tense mid vowels, indicating a coda-/l/-dependent allophone difference between the Netherlands and Flanders, most clearly for the (eː) vowel. In addition, all tense mid vowels are phonetically monophthongal in Flanders, which is not the case in the Netherlands. For the F2, significant tongue root retraction is observed for vowels preceding coda /l/, which, for front vowels, is larger in the Netherlands than in Flanders. Implications for the phonological and phonetic representations of the tense mid vowels and diphthongs and their regional spread will be discussed, as well as the key finding: that of the two sound changes investigated, one is Neogrammarian but the other is not.

[1] Zwaardemaker, H. & Eijkman, L. P. H. 1924. *Leerboek der Phonetiek. Inzonderheid met betrekking tot het Standaard-Nederlands*. Haarlem: De erven F. Bohn.

[2] Van de Velde, H. 1996. *Variatie en verandering in het gesproken Standaard-Nederlands*. Nijmegen: Katholieke Universiteit Nijmegen.

[3] Van de Velde, H., van Hout, R., & Gerritsen, M. 1997. Watching Dutch change: a real time study of variation and change in Standard Dutch pronunciation. *Journal of Sociolinguistics 1(3)*, 361–391.

[4] Stroop, J. 1998. *Poldernederlands. Waardoor het ABN verdwijnt*. Amsterdam: Bert Bakker.

[5] van der Harst, S. 2011. *The vowel space paradox: a sociophonetic study on Dutch*. Utrecht: LOT.

[6] Van de Velde, H. & van Hout, R. W. N. M. 2003. Diftongering in het Standaard-Nederlands. In: T. Koole, J. Nortier, & B. Tahitu (Eds.), *Artikelen van de vierde sociolinguïstische conferentie* (pp. 486–497). Delft: Eburon.

[7] Adank, P. M., van Hout, R. W. N. M., & Smits, R. 2004. An acoustic description of the vowels of Northern and Southern Standard Dutch. *The Journal of the Acoustical Society of America 116(3)*, 1729–1738.

[8] Adank, P. M., van Hout, R. W. N. M., & Van de Velde, H. 2007. An acoustic description of the vowels of Northern and Southern Standard Dutch II: regional varieties). *The Journal of the Acoustical Society of America 121(2)*, 1130–1141.

[9] Adank, P. M. 2003. *Vowel normalization: a perceptual-acoustic study of Dutch vowels*. Katholieke Universiteit Nijmegen.

[10] Van Son, R.J.J.H & Pols, L.C.W. 2001d. The IFA Corpus: A phonemically segmented Dutch "Open Source" Speech Database. *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam 24*, 15–26.

[11] van Reenen, P. & Jongkind, A. (2000). De vocalisering van de /l/ in het Standaard-Nederlands. *Taal en Tongval 52(1)*, 189–199.

[12] Boersma, P. 2001. Praat, a system for doing phonetics by computer. *Glot International* 5:9/10, 341-345.

[13] Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. 2013. Random effects structure for confirmatory hypothesis testing: keep it maximal. *Journal of Memory and Language 68(3)*, 255–278.

[14] Pinheiro, J. & Bates, D. 2000. *Mixed-effects models in S and S-PLUS*. Springer Science & Business Media.

Phonetic restructuring in the vowel systems of two Cavite Chabacano dialects

Marivic Lesho
University of Bremen

This study documents how vowels vary according to prosodic factors and differing levels of substrate influence in two dialects of Cavite Chabacano, a Spanish-lexified, Tagalog-substrate creole spoken in Cavite City, Philippines. Studies on creole phonology have rarely taken into account the importance of phonetic detail in investigating how creole phonological restructuring takes place [1]. However, as in other contact situations involving second language acquisition [2, 3], I argue that creole substrate influence is evident not only at the phonological level but also at the phonetic level.

Like Spanish, Cavite Chabacano has a 5-vowel system of /i e a o u/, phonologically restructured from the Old Tagalog 3-vowel system of /i a u/. Stressed /e/ and /o/ are usually realized as [e] and [o], as in Spanish; however, unstressed /e/ and /o/ are often raised to [i] and [u] [4, 5]. Previous phonological descriptions claim that this unstressed mid vowel raising occurs particularly in final position, especially in the San Roque district of Cavite City [4, 6].

To further investigate this dialectal variation and the issue of possible substrate and superstrate influence, a word list task was conducted with 21 Chabacano speakers from the San Roque district and 17 from the Caridad district. Duration, F1, and F2 measurements were taken for vowels in stressed and unstressed position as well as in phrase-final and nonfinal position ($n = 14{,}543$). Vowel dispersion was also calculated based on these formant measurements. The results were analyzed using linear mixed effects models, with word and speaker as random effects and stress, phrasal position, vowel category, and dialect as fixed effects. Tagalog substrate influence was expected in terms of unstressed vowel reduction as well as phrase-final prominence [7, 8], e.g. greater dispersion and duration in the final syllable, regardless of stress. Vowel category overlap [9] between /e/ and /i/ and between /o/ and /u/ in each prosodic condition was also calculated.

In both dialects, unstressed vowels were significantly shorter ($p < 0.0001$) and less dispersed ($p < 0.0001$) than stressed vowels. In addition, regardless of stress, vowel duration was longer in phrase-final than in non-final position ($p < 0.001$). The vowel dispersion model also showed that unstressed vowels had significantly greater dispersion in final than in nonfinal position ($p < 0.0001$), and there was a significant interaction between dialect and vowel category. San Roque had only a slightly less dispersed vowel space than Caridad overall ($p > 0.05$), but /e/ ($p < 0.0001$), /i/ ($p < 0.0001$), and /u/ ($p < 0.01$) were significantly less peripheral in that dialect.

Both dialects had substantial acoustic overlap between the high and mid vowels in unstressed position. For unstressed /i/ and /e/, however, San Roque had greater overlap (52%) than Caridad (35%), confirming previous descriptions of /e/-raising being more common there. However, the results were more similar for /u/ and /o/, with Caridad at 78% and San Roque at 75%. There is greater overlap between the two back vowels than the front vowels because in addition to unstressed /e/ and /o/ tending to be raised, /u/ is lower than /i/ in the vowel space.

Taken together, these results suggest that while both dialects have strong substrate influence (e.g., unstressed vowel reduction and phrase-final prominence), phonetic restructuring from the Old Tagalog 3-vowel system toward the Spanish 5-vowel system occurred to a slightly greater degree in Caridad than in San Roque, likely for reasons related to the sociohistorical settlement patterns of Cavite [10]. The vowels in Caridad are generally more dispersed, making room for a greater distinction between /i/ and /e/. These findings demonstrate that in situations where creoles have restructured by acquiring new phonological contrasts, substrate influence can still be found at the fine-grained phonetic level.

[1] Russell Webb, E. 2010. Creole phonological restructuring: The role of perception in contact-induced change. *Journal of Pidgin and Creole Languages*, 25(2), 263–288.
[2] Flege, J. 1995. Second language speech learning: Theory, findings, and problems. In Strange, W., (Ed.), *Speech perception and linguistic experience: Issues in cross-language research*. Timonium, MD: York Press, 233–277.
[3] Guion, S.-G. 2003. The vowel systems of Quichua-Spanish bilinguals. *Phonetica*, 60, 98–128.
[4] German, A.-B. 1932. *The Spanish dialect of Cavite*. Quezon City: University of the Philippines MA thesis.
[5] Ramos, F.-G. 1963. *A contrastive analysis of the sound system of the Cavite Chavacano and the English languages*. Manila: Abad Santos Educational Institution MA thesis.
[6] Miranda, G. 1956. *El dialecto chabacano de Cavite*. Unpublished manuscript.
[7] Gonzalez, A. 1973. Acoustic correlates of accent, rhythm, and intonation in Tagalog. *Phonetica*, 22, 11–44.
[8] Anderson, V.-B. 2006. Lexical stress without postlexical head-marking: Evidence from Tagalog. *Journal of the Acoustical Society of America*, 120(5), 3092–3092.
[9] Wassink, A.-B. 2006. A geometric representation of spectral and temporal vowel features: Quantification of vowel overlap in three linguistic varieties. *Journal of the Acoustical Society of America*, 119, 2334–2350.
[10] Lesho, M. 2013. *The sociophonetics and phonology of the Cavite Chabacano vowel system*. Columbus, OH: The Ohio State University dissertation.

Voice Onset Time in Brazilian Portuguese-Pomerano bilinguals

Mario Ruiz Moreno[1] and Christoph Gabriel[2]
[1]*Universität Hamburg,* [2]*Johannes Gutenberg-Universität Mainz*

Pomerano is a scarcely researched Low German variety taken to Brazil in the 1850s, which is still spoken by the descendants of the immigrants in the respective communities. Regarding its phonology, Pomerano is commonly described as being influenced by the long-lasting contact with Brazilian Portuguese (BP). Influence from BP shows up, among other things, in the pronunciation of the Pomerano front rounded vowels, which are replaced by their unrounded counterparts at least in some of the speakers (e.g. Pom. [m**y**ts] > [m**i**ts] 'cap' ([1, p. 52]). Multilingualism has also left its traces in the VOT patterns of stop consonants in the two languages spoken by the bilinguals (Voice Onset Time = time that elapses between the release burst and the vowel onset; see [2], [3]). Regarding VOT, the language pair under investigation displays the typical Germanic vs. Romance contrast, in that the phonological fortis-lenis (or voiceless-voiced) opposition is phonetically realized by means of a long vs. short lag distinction in Pomerano (i.e. aspirated [p^h t^h k^h] vs. unaspirated/devoiced [p t k]/[b̥ d̥ g̥]), whereas BP exhibits a voicing lead for the voiced stops and a short lag for their voiceless counterparts (i.e. pre-voiced [b d g] with negative VOTs vs. unaspirated [p t k]). The present contribution aims to determine the VOT values for both Pomerano and BP spoken in the Pomeranian community of Pomerode (Santa Catarina/Brazil) as well as to figure out whether older vs. younger bilinguals differ regarding the degree of phonic convergence shown in both languages. A further goal consists in providing a sociolinguistic interpretation of the results based on Labov's concept of covert prestige ([4]).

The data were collected in Pomerode in January 2015, using a picture naming task. 14 Pomerano-BP bilinguals and 4 monolingual speakers of BP were recruited and divided into three groups: (1) BP monolinguals (2♂, 2♀; average age: 39); (2a) younger bilinguals <70 (6♂, 1♀; average age: 57); (2b) older bilinguals >70 (4♂, 3♀; average age: 81). We recorded 18 test items in Pomerano and 19 in BP with voiceless/voiced stops in the onset of the stressed initial syllable, followed by both front and back vowels (e.g. Pom. *tung* 'tongue', *dek* 'blanket'; BP *pera* 'pear', *gota* 'drop'). VOT values were measured using Praat ([5]).

The results for **Pomerano** show for both groups of bilinguals (2a, 2b) the expected Germanic contrast (i.e. aspirated vs. devoiced/unaspirated stops); see Table 1. This finding largely confirms earlier studies carried out in regions other than Santa Catarina (see [6], [7] for the varieties of Pomerano spoken in Espírito Santo and Rio Grande do Sul). Regarding **BP**, the monolinguals (group 1) present the typical Romance contrast (i.e. voicing lead for /b d g/ vs. short lag for /p t k/), again as expected. Interestingly enough, the younger bilinguals (2a) display a mixed Romance-Germanic system in that they pre-voice /b d g/ (as the BP monolinguals do) on the one hand, but aspirate the voiceless stops on the other. The older bilinguals (2b), finally, use the same Germanic (i.e. unaspirated/devoiced vs. aspirated) contrast in both of their languages, thus showing transfer from Pomerano to BP; see Table 2.

As pointed out by [8], when a group is under attack, its speakers tend to stress what differentiates them for other groups. In this context, it is interesting to note that the older bilinguals from our sample (group 2b), who completely transfer their Germanic VOT values to BP, were brought up during Getúlio Vargas' dictatorship (1937–1945), a period characterized by repressions against the Germanic communities. Thus, covert prestige (see [4]), as becomes manifest in the stressing of a Germanic feature of pronunciation, still seems to operate within the older generation. By contrast, the younger bilinguals (group 2b), whose BP VOT values represent an intermediate system between Germanic and Romance (see Table 3), seem to attach higher priority to the (overt) prestige of a non-Germanic-accented pronunciation, at least with respect to their productions of BP voiced stops.

Pomerano	(2a) younger bilinguals		(2b) older bilinguals	
	VOT	SD	VOT	SD
/p/	43	15.49	51	13.49
/t/	51	12.69	56	8.28
/k/	71	11.75	71	11.83
/b/	14	3.06	14	4.90
/d/	25	9.95	21	7.91
/g/	no data *	--	no data *	--

Table 1. *VOT values (ms) and SD for Pomerano; differences between groups n.s.*
** No data obtained; initial* /ɡ/ *was produced as a velar approximant by all speakers.*

BP	(1) monolinguals		(2a) younger bilinguals		(2) older bilinguals	
	VOT	SD	VOT	SD	VOT	SD
/p/	15	1.47	27	6.32	38	14.00
/t/	21	1.97	31	1.60	37	5.40
/k/	46	8.29	49	9.62	51	11.07
/b/	-66	9.89	-76	24.76	-4	30.45
/d/	-60	11.74	-83	27.23	5	24.79
/g/	-58	20.60	-63	24.58	16	15.47

Table 2. *VOT values (ms) and standard deviation (SD) for BP; all differences statistically significant (Kruskal-Wallis, Mann-Whitney) between groups except for* /k/.

	BP			Pomerano		
	pre-voiced	**short lag**	**long lag**	**pre-voiced**	**short lag**	**long lag**
(1) monolinguals	[b d ɡ]	[p t k]				
(2a) younger bilinguals	[b d ɡ]		[pʰ tʰ kʰ]		[b̥ d̥ ɡ̥]	[pʰ tʰ kʰ]
(2b) older bilinguals		[p t k]	[pʰ tʰ kʰ]		[b̥ d̥ ɡ̥]	[pʰ tʰ kʰ]

Table 3. *Phonetic realization of the voiced-voiceless contrast in BP and Pomerano.*

[1] Schaeffer, S. C. B., & Meireles, A. R. 2014. Descrição sonora da língua pomerana. *Letras de Hoje* 49, 46–55.

[2] Lisker, L., & Abramson, A. S. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20, 384–422.

[3] Cho, T., & Ladefoged, P. 1999. Variation and universals in VOT. Evidence from 18 languages. *Journal of Phonetics* 27, 207–229.

[4] Labov, W. 1966/2006. *The social stratification of English in New York.* Cambridge: Cambridge University Press.

[5] Boersma, P., & Weenink, D. 2008. *Praat: doing phonetics by computer*. Computer program.

[6] Bandeira, M. H. T., & Zimmer, M. C. 2011. A transferência dos padrões de VOT de plosivas surdas no multilinguismo. *Letras de Hoje* 46, 87–95.

[7] Schaeffer, S. C. B., & Meireles, A. R. 2011. Padrões de vozeamento de consoantes plosivas em falantes de Pomerano (L1) e de Português (L2). *Anais do VII Congresso Internacional da Abralin*, 4009–4021.

[8] Trudgill, P. 2000. *Sociolinguistics: An introduction to language and society*. London: Penguin.

Acoustics of constricted epilarynx: The case of voiced coronal pharyngealized consonants in Jordanian and Moroccan Arabic

Jalal Al-Tamimi
Speech and Language Sciences, Newcastle University

Following the "Laryngeal Articulator Model" (LAM) [1], an epilaryngeal constriction causes tongue root/epiglottis retraction and voice quality changes [1-9]. The former leads to vowels produced with this configuration to be "retracted" with a back and down gesture causing the whole tongue to be retracted, including body and root [1-5]. The latter leads to a laryngealized or tense/pressed voice quality through ventricular folds constriction [6-7] and "enhanced" quality [8-9]. This exploratory study evaluates this type of constriction based on an acoustic analysis. A few articulatory reports suggest that pharyngealization in Arabic has retraction of the epiglottis, tongue body and root, as well as larynx raising that may lead to a tense/pressed voice quality [10-13]. Acoustically, pharyngealization in Arabic was shown to cause lowering of F2, raising of F1 and a general rising of F3 in the vowel [10-11]. Combined metrics may be better suited to show the retracted quality of these vowels.

This study aims at evaluating absolute and bark-difference formant frequencies as well as voice quality metrics to assess which acoustic parameters are indicative of this type of constriction in Arabic. The two dialects may show differences in their "guttural" quality. Twenty male speakers (10 per dialect) produced vowels preceded by /d or dˤ/. A total of 13 acoustic parameters measured at both onset and midpoint (total of 26) are obtained. These include: absolute formant frequencies (1-3), critical-band (Z) bark-difference (Z1-Z0, Z2-Z1, and Z3-Z2), spectral slope harmonic-differences of *f*0 and formant peaks (*H*1*-*H*2*, *H*1*-*A*1*, *H*1*-*A*2*, *H*1*-*A*3*, *A*1*-*A*2*) and the high frequency components (*A*1*-*A*3* and *A*2*-*A*3*). If pharyngealization is indeed associated with an epilaryngeal constriction, it is expected to observe lowering of F2 and an increase in Z3-Z2, an increase in F1, Z1-Z0, and with a more compacted spectrum through Z2-Z1. If spectral slope results are lower in the pharyngealized set, this would be indicative of a tense/pressed voice quality [14-15]. And finally, constricting the epilarynx on its own leads to an increase in the amplitude of harmonics in upper frequencies around F3 [8-9]. A total of 30966 measurements (17992 in JA and 12974 in MA) were statistically analysed. Given the high correlation between some of the predictors, we ran separate Bayesian Generalized Linear Mixed Effects modelling (GLMMs) on the z-scored predictors to allow for comparability. GLMMs were used to assess the predictive strength of each predictor. Non-significant predictors were eliminated through Likelihood ratio tests, and the standardized effect sizes were used. Then, an exploratory Random Forests classification technique was used to evaluate the classificatory power of these parameters through percent correct and degree of importance of these parameters. This technique is well suited to deal with correlated data, as we have here.

In both dialects, vowels in the vicinity of pharyngealized consonants are "retracted"; with a more open (higher F1, Z1-Z0) and more back (lower F2, higher Z3-Z2), with a more compacted spectrum (lower F2-F1), and with spectral divergence as an enhancing cue (higher Z3-Z2). Lower spectral slope results showed these vowels to have a more tense/pressed voice quality and with an increased energy in high frequencies; correlates of epilaryngeal constriction. Random forests results (Figure 1) on all acoustic cues showed extremely high classification rates (93.2% in JA, and 91.2% in MA), with the acoustic parameters varying in order of importance. Formant-based measures outperformed voice quality correlates leading to "retraction" being primary and voice quality secondary. Inter-dialectal variations are present with varying orders of parameters. These novel results indicate that Z2-Z1 may be the acoustic-perceptual correlate of "retracted" vowels in LAM, with tense/pressed voice and high frequency components being correlates of a constricted epilarynx, as used in Arabic.

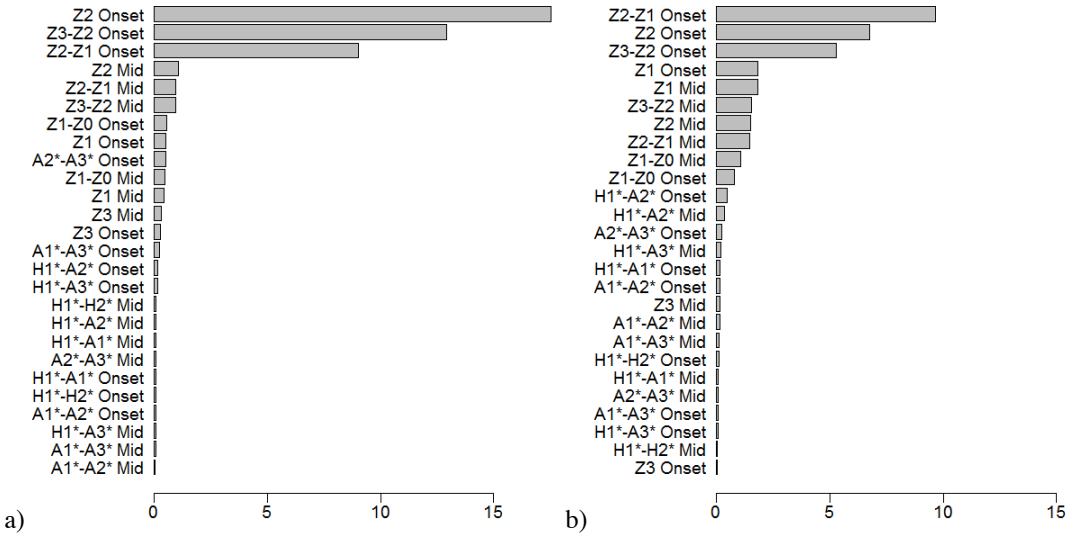


Figure 1: Mean percent decrease in accuracy importance scores of acoustic parameters in JA (a) and MA (b)

[1] Esling, J. H. (2005). There Are No Back Vowels: The Laryngeal Articulator Model. *The Canadian Journal of Linguistics / La revue canadienne de linguistique*, 50(1), 13-44

[2] Moisik, S. (2013). *The epilarynx in speech*. PhD thesis. Department of Linguistics, University of Victoria.

[3] Moisik, S., Czaykowska-Higgins, E., & Esling, J. (2012). The epilaryngeal articulator: a new conceptual tool for understanding lingual-laryngeal contrasts. *McGill Working Papers in Linguistics*, 22(1), 1-15.

[4] Sylak-Glassman, J. (2014a). *Deriving Natural Classes: The Phonology and Typology of Post-Velar Consonants* (Unpublished doctoral dissertation). University of California, Berkeley

[5] Sylak-Glassman, J. (2014b). An Emergent Approach to the Guttural Natural Class. *Proceedings of the Annual Meetings on Phonology*, 1-12

[6] Edmondson, J. A., & Esling, J. H. (2006). The valves of the throat and their functioning in tone, vocal register and stress: laryngoscopic case studies. *Phonology*, 23(2), 157–191

[7] Stevens, K. N. (1977). Physics of laryngeal behavior and larynx modes. *Phonetica*, 34(4), 264-279.

[8] Story, B. H. (2016, to appear). The Vocal Tract in Singing. In G. Welch, D. Howard, & J. Nix (Eds.), *The Oxford Handbook of Singing*, 1-21

[9] Titze, I. R., & Story, B. H. (1997). Acoustic interactions of the voice source with the lower vocal tract. *The Journal of the Acoustical Society of America*, 101(4), 2234-2243

[10] Al-Tamimi, F., & Heselwood, B. (2011). Nasoendoscopic, videofluoroscopic and acoustic study of plain and emphatic coronals in Jordanian Arabic. In B. Heselwood & Z. Hassan (Eds.), *Instrumental studies in Arabic phonetics* (p. 165-191). John Benjamins.

[11] Zeroual, C., & Clements, G. N. (2015). The feature [pharyngeal]. In A. Rialland, R. Ridouane, & H. van der Hulst (Eds.), *Features in Phonology and Phonetics: posthumous Writings by George N. Clements and Coauthors* (p. 247-276). Mouton de Gruyter: Berlin

[12] Laufer, A., & Baer, T. (1988). The emphatic and pharyngeal sounds in Hebrew and in Arabic. *Language and Speech*, 31, 181-205

[13] Lehn, W. (1963). Emphasis in Cairo Arabic. *Language*, 39(1), 29-39

[14] Garellek, M. (2012). The timing and sequencing of coarticulated non-modal phonation in English and White Hmong. *Journal of Phonetics*, 40(1), 152-161

[15] Kuang, J., & Keating, P. (2012). Glottal articulations of phonation contrasts and their acoustic and perceptual consequences. *UCLA Working Papers in Phonetics*, 111, 123-161.

Monday, 16:00 - 17:00
Poster Session 2, #3

Comparing coarticulatory directions in child speech

Elina Rubertus[1], Dzhuma Abakarova[1], Jan Ries[1], and Aude Noiray[1,2]
[1]*Laboratory for Oral Language Acquisition, University of Potsdam,* [2]*Haskins Laboratories*

In spoken language, phonemes which we think of as abstract phonological representations, are assembled into a continuous physical speech stream. This assembling is not just a pure concatenation of elements but results in context effects due to articulatory overlap: Coarticulation. In adults, it has been shown that coarticulatory effects are not only observed between adjacent segments but can span several segments in both the anticipatory and the carryover directions (for review [1]). Moreover, coarticulation effects in the two directions are claimed to originate from different underlying processes: While articulatory planning and preprogramming is the driving force for anticipatory coarticulation, carryover coarticulation originates from mechanical constraints and articulator inertia [2]. Coarticulatory processes have also been addressed in the framework of spoken language acquisition. However, most developmental studies on coarticulation focused on the anticipatory direction only (e.g., [3], [4], [5]). To further our understanding of how phonological knowledge is transposed into speech, which role a fine speech motor control plays, and which conclusions about linguistic representations and organization can be drawn from a change in coarticulatory magnitude and pattern across childhood, both coarticulatory directions need to be considered and compared.

The present study is the first to compare anticipatory and carryover V-to-V coarticulation in German children (3y, 4y, 5y, & 7y) and adults. With ultrasound imaging, tongue positions were directly traced instead of inferred from the acoustic signal. A symmetrical stimulus structure ($əC_1VC_2ə$) allowed us to test influences of the medial tense long vowel (/i/, /y/, /u/, /a/, /e/, /o/) on both schwas – the preceding one for anticipatory and the following one for carryover coarticulation. To investigate whether different mechanisms underlie the two directions, the intervocalic consonants varied in coarticulatory resistance (/d/>/g/>/b/). A highly resistant consonant is unlikely to change according to context while a low resistant consonant has less articulatory constraints and is thus more flexible to coarticulate. In line with Recasens, it was hypothesized that a resistant intervocalic consonant would decrease V-to-V coarticulation to a greater extent in the carryover (mechanical) than in the anticipatory (planning) direction [2].

Results indicate that children in all investigated age groups as well as adults exhibit measurable anticipatory as well as carryover V-to-V coarticulation. The magnitude of anticipatory coarticulation decreases with age, being strongest for three-year-olds and weakest for the adult participants. Interestingly, anticipatory V-to-V coarticulation is not affected by the resistance degree of the intervocalic consonant. In the carryover direction, the magnitude decrease across age is not as pronounced as in anticipatory coarticulation. However, here the intervocalic consonant does show influences on the amount of V-to-V coarticulation. Implications of this discrepancy for both the underlying mechanisms of the two processes as well as for the speech development of children will be discussed in more detail.

[1] Hardcastle, W. J. & Hewlett, N. 1999. *Coarticulation: Theory, Data and Techniques*. Cambridge: Cambridge University Press.

[2] Recasens, D. 1987. An acoustic analysis of V-to-C and V-to-V coarticulatory effects in Catalan and Spanish VCV sequences. *Journal of Phonetics*, *15*, 299-312.

[3] Nittrouer, S. 1993. The emergence of mature gestural patterns is not uniform: Evidence from an acoustic study. *Journal of Speech and Hearing Research, 36*, 959-972.

[4] Nittrouer, S. Studdert-Kennedy, M., & Neely, S. T. 1996. How children learn to organize their speech gestures: Further evidence from fricative-vowel syllables. *Journal of Speech and Hearing Research, 39*, 379-389.

[5] Nijland, L., Maassen, B., van der Meulen, S., Fabreels, F., Kraaimaat, F. W., & Schreuder, R. 2002. Coarticulation patterns in children with apraxia of speech. *Clinical Linguistics & Phonetics, 16*, 461-483.

A cross-linguistic perspective on stress-meter alignment in music: Evidence from Turkish children's songs

Barış Kabak and Christina Domene Moreno
University of Würzburg

Vocal music has been used to explore the ramifications of accentual prominence for the language user. In English and German, for instance, stressed syllables have been shown to align with strong beats of musical meter [1], suggesting the lexical nature of stress in these languages. In French, where stress is not lexical but fixed to the final syllable of an (accentual) phrase [2], metrical prominence in music has also been shown to align with word-final positions, suggesting the existence of word-final stress in French [3]. This research has been extended to pitch accent systems such as Tokyo Japanese (TJ), where accented syllables tend to fall on strong beats, with some degree of correspondence between tonal and melodic transitions [4]. The consequences of these findings for prosodic typology are unclear, however, since the outcome so far culminates on the isomorphism between linguistic accent and musical prominence despite employing languages with traditionally distinct accentual systems (English/German lexical stress, French: phrasal stress, TJ: pitch-accent system).

Here, we extend this research to Turkish, a mixed system, where prominence, regularly word-final, is modulated by lexical and morpho-syntactic properties, leading to non-final word stress [5]. Recent intonation studies classify Turkish as a pitch-accent system, with word-final stress as an artefact of prosodic phrasing [6], [7]. Studies exploring the significance of Turkish stress for the language user reveal an even more complex picture: In speech processing, listeners experience difficulties identifying stressed syllables [8] and evaluating lexical stress violations [9] while reliably using stress cues for word segmentation [10], [11].

Given the results from French and TJ, we also expect stress to manifest itself in Turkish music, at least in words with lexical stress. Following the methodology used in [3], we asked (i) whether there is a tendency for stressed syllables (final/non-final) in Turkish to coincide with strong musical positions, and (ii) whether the final syllables of polysyllabic words coincide with these positions even in non-phrase final contexts. We also asked which other phonological phenomena might modulate text-to-tune alignment in Turkish songs.

Our corpus consists of the lyrics of 24 Turkish children's songs composed in the "makam" (Corp1, n=12) and western traditions (Corp2, n=12). Each makam song employs one of five different "usuls" (rhythmic patterns, each with distinct metrical weight distribution, see [12]). The western songs were composed in 4/4 meter as in [1]. Syllables (n=2262) were coded for stress, syllable structure, and position in phrase, as well as for metrical weight (by inspecting the musical score). We modelled the data with ±stress, ±word initial position, and syllable-type as independent variables and song as a random effect in a linear mixed effects model.

Results differ for Corp1 and Corp2: In Corp1 neither stress nor syllable type or vocalic duration predict mean metrical weight (MW), thus indicating a lack of isomorphism between musical prominence and prosodic factors in the makam songs. However, mean MW in these songs was significantly higher for word-initial syllables, which thus align more strongly with musical prominence than any other syllable. In Corp2, on the other hand, mean MW was significantly higher for stressed than for unstressed syllables (Figure 1), showing results largely on a par with [1, 2, 3], all of which employed Western music. This indicates that accent-meter alignment is modulated by musical genre, and in genres where musical prominence is not isomorphic to accentual prominence, other types of alignment emerge, with word-edges serving as the target of the mapping between music and prosody. We suggest that matching meter with prosodic phenomena may be universal, but how alignment is manifested is genre-specific and gradient across languages.

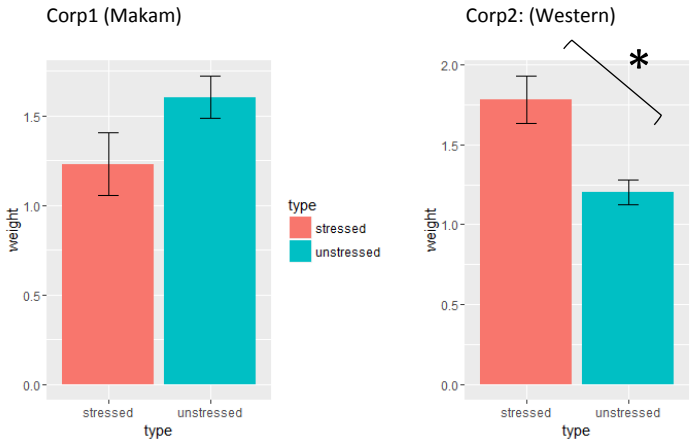


Figure 1. *Mean metrical weight for stressed vs. unstressed syllables in two musical genres.*

[1] Palmer, C. & Kelly, M. 1992. Linguistic prosody and musical meter in song. *Journal of Memory and Language* 31: 525-542.

[2] Jun, S. & Fougeron, C. 2002. Realizations of accentual phrase in French intonation. *Probus* 14: 147-172.

[3] Temperley, N. & Temperley, D. 2013. Stress-meter alignment in French vocal music. *Journal of the Acoustical Society of America* 134 (1) : 520-527.

[4] Cho, S. 2017. Text-setting in Japanese children's songs. *UPenn Working Papers in Linguistics*, 23.1.

[5] Kabak, B. & I. Vogel. (2001). The Phonological Word and Stress Assignment in Turkish. *Phonology* 18: 315-360.

[6] Kamali, B. 2011. *Topics at the PF interface of Turkish*. Doctoral Dissertation, Harvard University, USA.

[7] Güneş, G. 2015. *Deriving prosodic structures*. Doctoral Dissertation. University of Groningen, The Netherlands.

[8] Altmann, H. 2006. *The perception and production of second language stress: A cross-linguistic experimental study*. Doctoral Dissertation. University of Delaware, USA.

[9] Domahs, U., Genç, S., Knaus, J., Wiese, R. & B. Kabak. 2012. Processing (un)-predictable word stress: ERP evidence from Turkish. *Language and Cognitive Processes*, 1-20.

[10] Kabak, B., K. Maniwa, & N. Kazanina. 2010. Listeners use vowel harmony and word-final stress to spot nonsense words: A study of Turkish and French. *Laboratory Phonology* 1: 207-224.

[11] Van Ommen, S. 2016. *Listen to the beat: A cross-linguistic perspective on the use of stress in segmentation*. Doctoral Dissertation. Utrecht Institute of Linguistics, The Netherlands.

[12] Holzapfel, A. & Bozkurt, B. (2012). Metrical strength and contradiction in Turkish makam music. *Proceedings of the 2nd CompMusic Workshop*. Istanbul, Turkey, July 12- 13.

Monday, 16:00 - 17:00
Poster Session 2, #5

Dynamic Aspects of the Alveolar Fricative vs. Stop Contrast in Parkinson's Disease

Massimiliano M. Iraci, Mirko Grimaldi and Barbara Gili Fivela
CRIL-DREAM – Università del Salento

Dysarthria may differently affect speech articulators and the tongue seems to let report the highest variability of results. In Parkinson's Disease (PD) gestures by jaw, lower lip, and tongue, [1] found rarely reduced amplitude but slower tongue movements than jaw and lips', compared to healthy controls (HC). In PD's alveolar and velar plosives' release, [2, 3] found reduced duration but increased maximum velocity, acceleration, and deceleration. Differently from alveolars, this was found also during velars' closure. Results of our previous studies [4, 5, 6] showed that PD speakers' lip gestures amplitude in bilabial sequences could be reduced, and duration decreased compared to HC; but, in the same sequences, the tongue mobility was regularly increased in front-back dimension. In [7], PD alveolar fricatives (/s/ and /z/) were second only to stops (/k/ and /g/) in terms of perceptually individuated alterations: the misarticulation was described as a reduction of the tongue blade elevation towards the alveolar target in comparison with HC. The variability of results for tongue gestures may depend on the fact that the various parts of the tongue (e.g. root, dorsum, tip) are actively or passively involved in the realization of a number of speech gestures. Our goal is to check if tongue gestures by PDs vary depending 1) on the passive-to-active role of a specific effector, in particular the tongue tip being passive for a vocalic and active for a specific consonantal gesture; 2) on the gesture to be realized, as in the case of the tongue tip (realizing a closure vs. a narrowing). Our hypotheses are that both a) passive-to-active and b) specific gestures to be realized have a specific impact on PDs' in comparison to HC's productions.

Tongue dynamics in alveolar stops and fricatives are compared in a stable vowel context (/a/-to-/a/) in order to observe the tongue motor control in sequences of Cs and Vs controlled, respectively, by tongue tip and dorsum. Moreover, fricatives require a subtler motor control for the tongue than stops: the articulator cannot hit and rest on a target – thus using it to help controlling the release – but has to stop at the correct distance as to produce the typical turbulent sound. Acoustic and kinematic data from 5 Italian mild-to-severe PD dysarthric speakers (age: 71,8, s.d.±7,19) and age-matched HCs (age: 70,6, s.d.±7,02) have been recorded through Electromagnetic Articulography (AG501). Target items contained unvoiced alveolar stops and fricatives (/t/ vs. /s/). Dynamical parameters were measured and analyzed for the purpose of potential changes in the motor behavior [8]. Results show that HCs realize the difference between stops and fricatives similarly to an expected change in the target vertical position [see 8]: no duration differences are found between the two gestures, but stops are regularly realized with higher peak velocity, amplitude, and stiffness. Surprisingly, PD speakers, on average, seem to behave similarly to HCs in the production of fricatives, but differently in that of stops: during closure, peak velocity and stiffness are higher in stops than in fricatives, while duration decreases; in the release, duration, amplitude and peak velocity increase in stops, and stiffness does not in comparison with fricatives.

Results suggest that PDs realize fricatives similarly to HCs, exerting an effort to guarantee the realization of a narrowing rather than a closure. As for stops, the closure greater stiffness and lower duration could suggest that PDs do not finely control the tongue tip, but rather exploit a physical obstacle (i.e. alveolar ridge), hitting it to end the gesture. This way, they create the conditions for a longer release, as if a higher than expected target were reached and a longer opening gesture needed (as in the case of long segments [9]). Individual strategies will be discussed. However, data suggest that PDs a) do not necessarily have troubles in switching from a passive-to-active role of the tongue tip (e.g., from passive in [a] to active in [s]). Rather, it is b) the switch plus a specific gesture (hitting on a physical obstacle) that seems to cause a change in their articulatory strategy.

		Amplitude		Duration		Stiffness		Peak Velocity	
		Closure	Release	Closure	Release	Closure	Release	Closure	Release
HC	/t/	9.06	8.04	132.96	126.90	21.43	19.43	178.20	148.37
	/s/	7.85	6.60	137.75	132.89	15.88	15.85	115.32	101.59
PD	/t/	8.74	8.73	103.62	143.41	19.97	14.83	167.67	125.73
	/s/	8.10	7.12	124.15	127.69	15.24	15.88	118.61	109.84

Table 1 - *Overall mean values: amplitude in millimetres, duration in milliseconds, peak velocity in millimetres/seconds*

Consonant closure					Consonant Release			
Amp	Dur	Stiff	PeakVel		Amp	Dur	Stiff	PeakVel
Fri<Stop 1.35(0.5)	/	Fri<Stop 5.63(0.7)	Fri<Stop 64.7(8.6)	**HC**	Fri<Stop 1.41(0.4)	/	Fri<Stop 3.6(1.2)	Fri<Stop 47.0(12.4)
/	Stop<Fri 7.56(2.1)	Fri<Stop 4.48(1.5)	Fri<Stop 54.2(17.4)	**PD**	Fri<Stop 1.00(0.6)	Fri<Stop 31.0(7.2)	/	Fri<Stop 18.9(6.8)

Table 2 - *Results of statistical tests (mixed models): significant results are reported with estimate and standard error*

[1] Yunusova, Y., Weismer, G., Westbury, J.R., & Lindstrom, M.J. 2008. Articulatory movements during vowels in speakers with dysarthria and healthy control. *Journal of Speech, Language, and Hearing Research*, 51 (3), 596-611.

[2] Wong, M.N., Murdoch, B.E., & Whelan, B.-M. 2010. Kinematic analysis of lingual function in dysarthric speakers with Parkinson's Disease: an electromagnetic articulography study. *International Journal of Speech-Language Pathology*, 12(5), 414– 425.

[3] Wong, M.N., Murdoch, B.E., & Whelan, B.-M. 2011. Lingual Kinematics in Dysarthric and Nondysarthric Speakers with Parkinson's Disease. *Parkinson's Disease*, 1-8.

[4] Gili Fivela, B., Iraci, M.M.., Sallustio, V., Grimaldi, M., Zmarich, C., & Patrocinio, D. 2014. Italian Vowel and Consonant (co)articualtion in Parkinson's Disease: extreme or reduced variability? *Proceedings of the 10th International Seminar on Speech Production* (Cologne), 146–149.

[5] Gili Fivela, B., Iraci, M.M., Grimaldi, M., & Zmarich, C. 2015. Consonanti scempie e geminate nel Morbo di Parkinson: la produzione di bilabiali. In Avesani, C., Vayra, M. & Tamburini, F. (Eds.), *Il Farsi e il disfarsi del linguaggio. Acquisizione, mutamento e destrutturazione della struttura sonora del linguaggio / Language acquisition and language loss. Acquisition, change and disorders of the language sound structure*. Milano: Studi AISV.

[6] Iraci, M.M., Zmarich, C., Grimaldi, M., & Gili Fivela, B. 2016. Il parlato nel Morbo di Parkinson: ampiezza dei gesti articolatori e distintività dei suoni linguistici. In Sorianello, P. (Ed.) *Il linguaggio disturbato. Modelli, strumenti, dati empirici*. Roma: Aracne.

[7] Logemann, J.A., & Fisher, H.B. 1981. Vocal tract control in Parkinson's Disease : Phonetic feature analysis of misarticulations. *Journal of Speech and Hearing Disorders*, 46, 348-352.

[8] Cho, T. 2006. Manifestation of Prosodic Structure in Articulation: Evidence from Lip Kinematics in English. In Goldstein, L., Whalen, D., & Best, C. (Eds.), *Laboratory Phonology 8*. Berlin: Mouton de Gruyter, 519-548.

[9] Löfqvist, A. 2005. Lip kinematics in long and short stop and fricative consonants. *Journal of the Acoustical Society of America*, 117, 2, 858-878.

Monday, 16:00 - 17:00
Poster Session 2, #6

Interaction between Word Length and Emotional Prosody

Seung Kyung Kim
Aix Marseille Univ, CNRS, LPL, Aix-en-Provence, France

A growing body of research shows that phonetically-cued social information in speech (e.g., speaker's age) can influence speech perception and spoken language understanding (e.g., Drager 2010). One fundamental question in the field is to understand and explain which principles and mechanisms underlie when listeners decode and integrate linguistic and social information in speech (Sumner et al. 2014). Contributing to this question, the current study investigates how strength of lexical information interacts with the effect of phonetically-cued social information in spoken word recognition. Specifically, I test whether word length modulates the effect of phonetically-cued *emotional* information (i.e., emotional prosody) and show that the effect emerges stronger when the prosody is carried by shorter words.

Previous work has shown that emotional prosody activates words associated with the emotion (Kim 2015, Kim & Sumner 2016). For example, hearing a non-emotional word (*pineapple*) uttered with angry prosody can facilitate recognition of angry-associated words (*mad, upset*). Building on this finding, I investigate whether word length modulates affective/prosodic priming between emotional prosody and emotional words. Word length is an important dimension in word recognition, as longer words produce stronger lexical activation than shorter words (Pitt and Samuel 2006). This is because longer words provide a larger amount of bottom-up acoustic signals than shorter words and because longer words tend to have less top-down inhibition from similar words than shorter words. My hypothesis is that phonetically-cued social information shows a stronger effect in linguistic processing when linguistic information is weaker. If this is true, we should find a stronger effect of emotional prosody in facilitating recognition of emotional words when the words carrying emotional prosody are shorter rather than longer. That is, a two-syllable word (*atom*) uttered in angry prosody should be more effective in facilitating recognition of angry-related words (*mad, upset*) than a four-syllable word (*aluminum*) in angry prosody does.

I test this prediction with a cross-modal lexical decision task. The visual target words were 12 angry-related words (e.g., *mad, upset*). The target was preceded by 48 two-, three-, or four-syllable non-emotional prime words (e.g., *atom, envelope, aluminum*) spoken with emotional prosody or with neutral prosody. The reaction times (Figure 1) were analyzed using linear mixed-effect models. The interaction between syllable length and prosody type reached significance (beta=0.02, se=0.01, p=0.05). When the prime words had 2 syllables, listeners (n=372) recognized angry target words faster after hearing prime words in angry prosody (585ms) than after hearing them in neutral prosody (595ms) (beta=0.02, se=0.009, p=0.05); when the prime words had 3 or 4 syllables, however, angry target words were not recognized faster after hearing angry prosody (590ms) than after hearing neutral prosody (586ms) (beta=-0.004, se=0.006, p=0.5).

This result confirms the hypothesis that word length modulates the effect of emotional prosody on spoken word recognition. The finding suggests that the strength of linguistic information is one crucial factor that interacts with social effects in spoken language processing, providing a critical insight into how linguistic and social information in speech interact in principled ways early in the process of understanding spoken language.

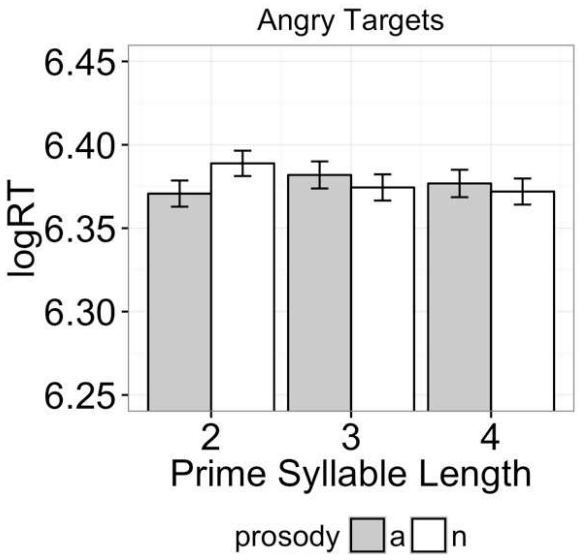


Figure 1. *Reaction time results*

[1] Drager, K. 2010. Sociophonetic variation in speech perception. *Language and Linguistics Compass*, *4*(7), 473–480.

[2] Kim, S. K. 2015. *Speech, variation, and meaning: The effects of emotional prosody on word recognition* (Doctoral dissertation). Stanford University, CA, USA.

[3] Kim, S. K. & Sumner, M. 2016. *Beyond lexical meaning: The effect of emotional prosody on spoken word recognition*. Manuscript submitted for publication.

[4] Pitt, M. A., & Samuel, A. G. 2006. Word length and lexical activation: Longer is better. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(5), 1120–1135.

[5] Sumner, M., Kim, S. K., King, E., & McGowan, K. B. 2014. The socially weighted encoding of spoken words: A dual-route approach to speech perception. *Frontiers in Psychology*, *4*, 1–13.

Monday, 16:00 - 17:00
Poster Session 2, #7

Gestures copying the F0 curve of lexical tones help learning Mandarin tones and words

Florence Baills[1], Nerea Suárez-González, Santiago González-Fuente[1], Pilar Prieto[1,2]
[1]*Department of Translation and Language Sciences, Universitat Pompeu Fabra (Spain),*
[2]*Institució Catalana de Recerca i Estudis Avançats (ICREA)*

Research in second language acquisition has described the positive effects of observing iconic gestures on vocabulary acquisition (e.g. Kelly et al., 2009; Macedonia et al., 2011; Tellier, 2008) as well as the positive effect of beat gestures on L2 pronunciation learning (Gluhareva & Prieto, 2016) and vocabulary acquisition (Kushch et al., under review). However, little is known about the role of a specific type of metaphoric gesture that mimicks melody in speech (e.g. the so-called pitch gestures, Morett & Chang, 2014) in the learning of L2 intonational variations across languages.

Morett & Chang (2014) reported that a brief training with pitch gesture production could have beneficial effects on Mandarin L2 word learning. Chen (2013) also reported better tonal production and a higher frequency of accurate responses by learners of Mandarin Chinese who were trained in the classroom to perceive and produce these gestures than learners following the traditional numeric system of tones. However, this classroom study did not control the training paradigm strictly – for example, there was more emphasis on tone learning in the gesture group than in the control group – nor did it address the potential beneficial effects of respectively observing or producing these gestures in the classroom.

In a between-subject training study, we explore the gains of observing (Experiment 1) and producing (Experiment 2) pitch gestures during a short multimodal training about Mandarin Chinese tones and words (a) on tones identification abilities and (b) on vocabulary learning. Tone identification abilities were trained with 18 minimal pairs of monosyllabic words and tested with a tone perception task. Vocabulary learning included twelve target words which were all tested after the training: in a word-meaning recall task, participants listened to the Chinese word and wrote down the meaning in Catalan. Then in an identification task, they listened to the Chinese word and had to choose between the meanings of two words from the same minimal pair. Participants sat in front of a computer in a silent room and wear headphones to watch the training in a video and subsequently carried out perceptive and vocabulary tasks. They weren't given any feedback during the tasks.

Experiment 1 (50 participants) tested the gains of observing vs. not observing pitch gestures. Results from three GLMM statistical analyses indicate that observing pitch gestures significantly strengthened (1) lexical tones identification; (2) word-meaning recall and (3) word-meaning association of newly presented Mandarin words that only differ in tone ($p<.05$ for (1) , $p<.001$ for (2) and (3)). Experiment 2 (57 participants) tested with the same tasks the gains of observing and vocally repeating vs. observing, vocally repeating and producing pitch gestures on the same abilities. Results from three GLMM statistical analyses show that producing gestures significantly favored tone identification and word recall more than merely repeating the word and viewing the gesture ($p<.01$ for (1), $p<.001$ for (2), $p<.05$ for (3)). Crucially, by comparing results from the two experiments, we found that gesture observation was significantly more beneficial to lexical tone identification ($p<.05$) and word-meaning association ($p<.05$) than gesture production.

Pitch gestures may or not already be used by CSL (Chinese as a Second Language) instructors to teach lexical tones; in both cases our findings can be an incentive to start implementing this multimodal approach or to adopt a more efficient methodology.

Chen, C.-M. (2013). Gestures as tone markers in multilingual communication. In *Research in Chinese as a Second Language*, pp. 143–168.

Gluhareva, D., & Prieto, P. (2016). Training with rhythmic beat gestures benefits L2 pronunciation in discourse-demanding situations. *Language Teaching Research*. JOUR.

Kelly, S. D., McDevitt, T., & Esch, M. (2009). Brief training with co-speech gesture lends a hand to word learning in a foreign language. *Language and Cognitive Processes*, *24*(2), 313–334.

Kushch, O., Igualada, A., & Prieto, P. (2016). *Gestural and prosodic prominence favor second language novel word acquisition*. Manuscript submitted for publication.

Macedonia, M., Müller, K., & Friederici, A. D. (2011). The impact of iconic gestures on foreign language word learning and its neural substrate. *Human Brain Mapping*, *32*(6), 982–998.

Morett, L. M., & Chang, L.-Y. (2014). Emphasising sound and meaning: pitch gestures enhance Mandarin lexical tone acquisition. *Language, Cognition and Neuroscience*, (August), 1–7.

Tellier, M. (2008). The effect of gestures on second language memorisation by young children. *Gesture*, 8, 219–235.

Monday, 16:00 - 17:00
Poster Session 2, #8

Phonological knowledge and phonetic details in the production of stop sequences by Mandarin Chinese native speakers

Qianwen Guan
Université Paris Diderot

Language experience leads to the emergence of epenthetic vowels as a very common repair strategy for illicit consonant sequences (CC) [3]. At the same time, epenthesis rate is affected by various phonetic factors, for instance, place of articulation of the consonants [7].

This study aims to assess the role of phonological knowledge and phonetic details in the production of non-native phonotactics by investigating the quality of the epenthetic vowel. Loanword studies generally show that epenthetic vowels in consonant sequences are consistent with native phonology [1] [6]. However, more detailed acoustic study shows that English speakers produce intrusive material within a phonotactically non-native CC sequence with shorter duration than English lexical vowels [2]. Intrusive material is transitional element which may be present acoustically but not phonologically. This is attributed to gestural mistiming: speakers fail to coordinate the consonants of the cluster and they produce a transitional release between two consonantal constrictions [4].

In the present study, epenthetic vowels are studied in a production experiment conducted with monolingual Mandarin Chinese (MC) speakers. MC is a language with simple phonotactics, not allowing stop-stop sequences. 24 native MC speakers were recorded in Beijing. The stimuli were 14 pseudo-words in Russian, containing CC sequences and controls with the vowel [a] between two consonants CaC, with two positions of the cluster relative to the stressed syllable: initial pre-stress (/ákta/) and medial post-stress (/ktápa/). The clusters included /kt/, /tk/, /tp/, /pt/. The words were excised from a carrier phrase in which they were originally produced by the Russian speaker. MC speakers repeated each word they heard.

Previous perceptual study from the same participants supports the role of native phonotactics [5]. The vowel used most frequently to transcribe the epenthetic vowel by MC speakers is a back mid unrounded vowel [ɤ] after both C1-[k] and C1-[t], but not after C1-[p]. Participants inserted a rounded vowel [u] after C1-[p], consistent with a co-occurrence restriction in MC phonology (*[pɤ]) and C1 place. Based on these findings, we predict that if phonological knowledge plays a role, the quality of the epenthetic vowel will be different after different C1 according to phonotactic restrictions.

Preliminary results from 10 participants are presented in this abstract. An epenthetic vowel was produced 47% of the time. The vowel was identified by the presence of voice bar, of formants and of periodic waveform between two consonants (Fig. 1). The ***duration*** of the epenthetic vowel produced in CC sequences is significantly shorter than that of the lexical vowel in controls CaC [β = -39.8, $p < 0.001$] (Fig. 2). Regarding the ***quality*** of the epenthetic vowel**,** F1 and F2 were measured at mid-point. Results indicated that there is no significant difference for F1 and F2 values after different C1 ([k], [t], [p]) [$p > 0.05$] (Fig. 3). The mean formant values in Table 1 shows that the epenthetic vowel after different C1 produced by MC speakers is a schwa-like material. Even though participants perceived a vowel [u] after C1-[p] in transcription [5], they did not really produce [u] acoustically. Moreover, there is no significant effect of position for either F1 or F2 values [$p > 0.05$].

Taken together, in the production of non-native consonant clusters, the presence of epenthetic vowels could be considered consequences of phonotactic restrictions, but the quality of epenthetic vowel could not be determined by native phonological knowledge.

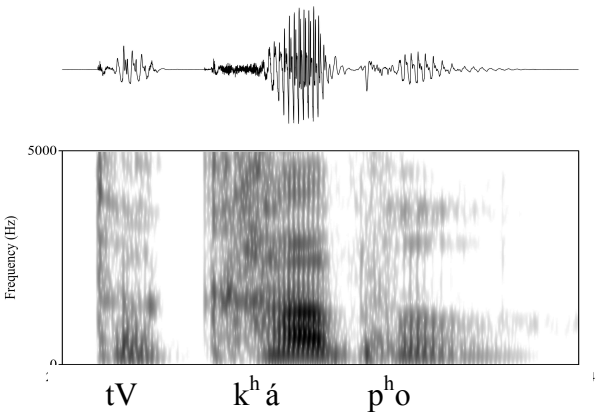


Figure 1. *Spectrograms illustrating the production of the Russian nonword with CC sequence /tkapa/ by a MC speaker.*

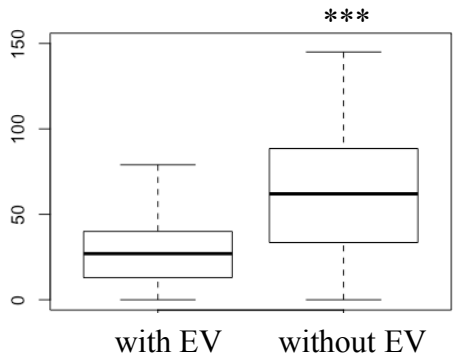


Figure 2. *Duration of C1 with epenthetic vowel (EV) and without epenthetic vowel.*

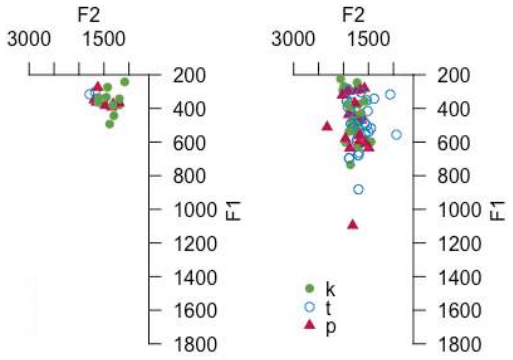


Figure 3. *F1 and F2 mid-point values for different vowels after k, t, p (kV, tV, pV) (male left vs. female right)*

C1	F1 (Hz)		F2 (Hz)	
	mean	s.d.	mean	s.d.
Female				
k	458	148	1729	256
p	493	186	1782	189
t	495	145	1675	243
Male				
k	357	36	1489	184
p	334	28	1631	127
t	359	73	1359	177

Table 1. *Mean formant values (F1 and F2) of epenthetic vowel by gender.*

[1] Broselow, E. 1999. Stress, epenthesis, and segment transformation in Selayarese loans. *Proceedings of the Annual Meeting*, Berkeley Linguistics Society 25. 311–325.

[2] Davidson, L. 2006. Phonotactics and articulatory coordination interact in phonology: Evidence from non-native production. *Cognitive Science*, 30(5), 837–862.

[3] Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., Mehler, J. 1999. Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance* 25, 1568-1578.

[4] Gafos, A. 2002. A grammar of gestural coordination. *Natural Language and Linguistic Theory*, 20(2), 269–337.

[5] Guan, Q., & Kwon, H. 2016. Phonotactic and prosodic adaptation of non-native consonant sequences by Mandarin native speakers. Poster presented at 5th Joint Meeting of the Acoustical Society of America and Acoustical Society of Japan, Hawaii

[6] Paradis, C., & Lacharité, D. 1997. Preservation and minimality in loanword adaptation. *Journal of Linguistics* 33. 379–430.

[7] Wilson, C., Davidson, L., Martin, S. 2014. Effects of acoustic–phonetic detail on cross-language speech production. *Journal of Memory and Language* 77, 1-24.

Monday, 16:00 - 17:00
Poster Session 2, #9

Synchronization of speech rhythm between Spanish-speaking interlocutors

Leonardo Barón Birchenall, Noël Nguyen
Aix Marseille Univ, CNRS, LPL, Aix-en-Provence, France

In spoken language communication, accommodation refers to the many processes employed by talkers to adapt to each other. Within these processes, phonetic convergence is associated with an increase in similarity in speech patterns between interlocutors over the course of a conversation. In some theoretical frameworks, phonetic convergence is thought to occur in an involuntary and automatic manner rather than intentionally [1].

The present work focuses on the interaction between speech rhythm and phonetic convergence in a semi-interactive task (understanding rhythm as a "temporally regular [iteration] of events which embody alternating strong and weak values of an observable parameter" [2]). Specifically, given that a repeated speech stimulus requires both less processing time and lower neural activation across repetitions, and that multiple repetitions significantly enhance memory and learning [3], we propose that the use of regular rhythmic structures during conversations produces more convergence between speakers with respect to irregular rhythmic structures. To our best knowledge, the only existing research on this particular topic is the one conducted by Späth et al. [4], who found more rhythmic convergence between a healthy person and a model speaker than between individuals with Parkinson's disease and the same model speaker.

To test our hypothesis, we created a set of stimuli consisting of seven groups of 16 nine- or eight-syllable Spanish sentences each. Each group had a particular rhythmic structure, obtained through the arrangement of different types of words (oxytones, paroxytones, proparoxytones and unstressed words) in feet of different length. Rhythmic structures were composed as follows (unstressed syllables are represented by a lowercase **x** and stressed syllables by an uppercase **X** and in uppercase within the sentences):

Regular structures: (1) Three feet, head to the right: **xxXxxXxxX** (e.g. la re-PÚ-bli-ca NO ter-mi-NÓ) [the republic did not end]. (2) Three feet, head to the left: **XxxXxxXxx** (e.g. E-llos es-PE-ran al MÉ-di-co) [they wait for the doctor]. (3) Four feet, head to the left: **XxXxXxXx** (e.g. JUAN es-TÁ bus-CAN-do BA-rro) [John is looking for mud].

Irregular structures: (4) Three feet, head to the right: **xxXxXxxXx** (e.g. la es-PO-sa CA-e sin VI-da) [the wife falls down dead]. (5) Three feet, head to the left: **XxxxXxxXx** (e.g. CAR-los ter-mi-NÓ mi lla-MA-da) [Charles ended my call]. (6) Four feet, head to the left: **XxXXxxxXx** (e.g. JUAN es-TÁ SIEM-pre so-me-TI-do) [John is always under control]. (7) Four feet, head to the left: **XxXXxxXx** (e.g. JUAN sa-LIÓ RÁ-pi-do SIEM-pre) [John always left quickly].

We tested four dyads of Spanish native speakers separately in a reading - repetition task with different combinations of the rhythmic structures. In the task, each member of the dyad must read a sentence and the other one must immediately repeat it. Participants alternate between reading and repeating the sentences of each group. The order of presentation of the sentences within a group, and that of the groups themselves, are randomized. A rhythmic distance score, proposed by Späth et al. [4], was then used to determine the degree of convergence between the interlocutors' rhythms.

Results indicate a greater amount of convergence between regular structures than between irregular ones, when feet nuclei are left aligned. We observed an overall tendency for the regular utterances to present more similar metrical timing patterns between interlocutors than the irregular ones, rather than a gradual augmentation of the resemblance between regular utterances' rhythms over the course of the task. Details will be given on the response patterns observed in the other conditions (right-aligned feet nuclei), and implications for current models of phonetic convergence in speech will be discussed.

[1] Louwerse, M., Dale, R., Bard, E., & Jeuniaux, P. 2012. Behavior matching in multimodal communication is synchronized. *Cognitive Science*, 36(8), 1404-1426.
[2] Gibbon, D. 2015. Speech rhythms – modeling the groove. In R. Vogel & R. van de Vijver (Eds.), *Rhythm in Cognition and Grammar: A Germanic Perspective* (pp. 108-161). Berlin: De Gruyter Mouton.
[3] Falk, S., Rathcke, T., & Dalla Bella, S. 2014. When speech sounds like music. *Journal of Experimental Psychology: Human Perception and Performance*, 40(4), 1491-1506.
[4] Späth, M., Aichert, I., Ceballos, A., Wagner, E., Miller, N., & Ziegler, W. 2016. Entraining with another person's speech rhythm: Evidence from healthy speakers and individuals with Parkinson's disease. *Clinical Linguistics and Phonetics*, 30(1), 68-85.

Monday, 16:00 - 17:00
Poster Session 2, #10

Diatopic variation in the prosodic realization of left-dislocations in Spanish

Ingo Feldhausen[1] and Alina Lausecker[1]
[1]*Goethe-Universität Frankfurt (Germany)*

This paper deals with the prosodic realization of clitic left-dislocations (CLLD) in European and Latin-American Spanish. By drawing from data of different production studies we first highlight the diatopic differences in the prosody of these constructions. After that, we propose a constraint-based approach in order to account for the attested dialectal variation.

The main syntactic characteristic of CLLDs is the presence of a phrase in the left periphery of a clause (e.g. *el libro* ‚the book' in (1a)), which is typically connected with the clause by means of an anaphoric element (the weak pronoun *lo* ‚it' in (1a)). Despite several studies on the syntax and information structure of CLLD construction in Spanish (e.g. Contreras, 1976; Hernanz & Brucart, 1987; Escobar, 1995; Casielles-Suárez, 2003; López, 2009; Olarrea, 2012; Zubizarreta, 1999), there are only a few detailed studies on their prosody (Feldhausen 2014: ch. 4.2, 2016a, b, Pešková 2015). In combination with the latter studies, we show which prosodic differences occur between CLLD constructions in Peninsular Spanish, *porteño*-Spanish, and Peruvian Spanish.

In previous research, the CLLD constituent is typically considered to from a single prosodic unit at the level of the Intonational Phrase (IP; e.g. Nespor & Vogel 1986, Frascarelli 2000, Féry 2011). Our data, however, suggests a finer grained picture, which is strongly shaped by the diatopic dimension. We conducted production experiments based on scripted speech with native speakers of all three aforementioned varieties of Spanish (Spain: N=8, Argentina: N=25, Peru: N=4). We investigate non-embedded (1a) and embedded (1b) CLLDs as well as simple (1a) and multiple (1b) CLLDs. The target sentences were introduced by context questions to control for the givenness of the CLLD constituent (e.g. López, 2009).

The results show that there is an obligatory prosodic boundary at the *right edge* of the CLLD element (both in simple and multiple CLLDs as well as in both embedded and non-embedded contexts) in Peninsular and Peruvian Spanish, see (2). The boundary can be stronger or weaker (i.e. Intonational Phrase (IP) or intermediate phrase (ip)) and can be accompanied by a pause. In *porteño*-Spanish, the right boundary can be considered to be typical, but not obligatory due to a lesser occurence frequency, which is between only 67% and 80%. In all varieties, the right edge is realized by a high boundary tone (H- or H%), see Figure 1. As for the *left edge* of CLLD constituents, embedded contexts are relevant, because only in such configurations the left-dislocated element is preceded by further material, namely the matrix clause (see *Pedro dice* 'Peter says…' in (1b)). In Peninsular Spanisch, the matrix clause is also obligatorily separated from the embedded clause containing the CLLD by a prosodic boundary. In Peruvian Spanish, the left boundary seems to be optional, since many instances of the CLLD constituent create a prosodic unit together with the matrix clause (around 50%). As a consequence, a clear mismatch between prosodic (2b) and syntactic (2c) structure occurs (similar phrasing patterns were also attested for Catalan (Feldhausen, 2010) and Bantu languages (Downing, 2011)).

Our data suggest that there may be an important difference between geographical varieties, confirming the important role of dialectal variation in intonational phonology (Prieto & Roseano, 2010; Vanrell & Fernández-Soriano, 2013; Feldhausen & Vanrell, 2015; Frota & Prieto, 2015). More concretely, our data show that the information structural homogenous category CLLD is prosodically not realized in a uniform way. This has consequences for approaches assuming a 1:1 relation between prosody and information structure. With Feldhausen (2016a) and Patin et al. (2017) we assume that the prosodic variation can be accounted for by the interaction of several violable constraints, among them ALIGN-TOP,R and ALIGN-CP,L. The former constraint requires a prosodic boundary at the right edge of CLLD constituents, while the latter demands a boundary at the left edge of the embedded clause (CP).

A different ranking on a continuous ranking scale (Boersma & Hayes, 2001) explains the inter- and intra-dialectal variation.

(1) a. **El libro**, *lo* ha dado a Ana. Non-embedded, simple CLLD
'The book, s/he gave it to Ana.'

b. Pedro dice que **el libro, a Ana**, *se lo* ha dado. Embedded, multiple CLLD
'Peter said that the book, to Ana, s/he gave it.'

(2) a. ( )( )( ) prosodic structure (Peninsular Spanish)
b. ( )( ) prosodic structure (Peruvian Spanish)
c. [CP1 [CP2]] syntactic structure
Pedro dice que **el libro,** se *lo* ha dado a Ana.

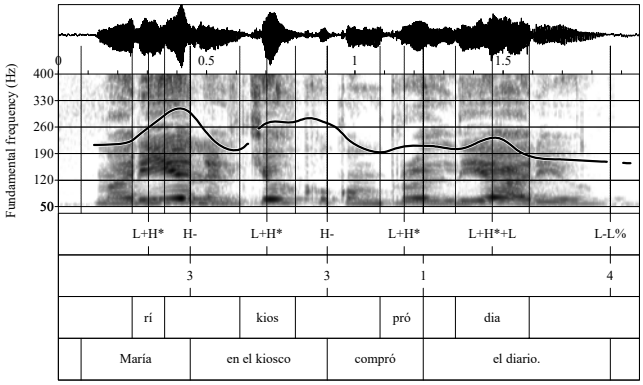


Figure 1. *Waveform, spectrogram, and F0 trace for the* porteño-*Spanish sentence* María en el kiosco compró el diario *'Mary bought the newspaper at the kiosk' with a H- edge tone after both dislocated constituents,* María *'Maria' and* en el kiosco *'at the kiosk' (which are not doubled by a clitic due to a defective clitic system).*

Selected references:

[1] **Boersma**, P. **& Hayes**, B. 2001. Empirical tests of the gradual learning algorithm. *Linguistic Inquiry* 32: 45–86.

[2] **Contreras**, H. 1976. *A Theory of Word Order with Special Reference to Spanish*. North-Holland.

[3] **Downing**, L. 2011. The Prosody of "Dislocation" in Selected Bantu Languages. *Lingua* 121(5): 772–786.

[4] **Feldhausen**, I. 2016a. Inter-speaker variation, Optimality theory, and the prosody of clitic left-dislocations in Spanish. *Probus* 28(2): 293-334.

[5] **Feldhausen**, I. 2016b. The Relation between Prosody and Syntax: The case of different types of Left-Dislocations in Spanish. In: Armstrong, M., N. Henriksen & M. Vanrell (Eds.). *Intonational grammar in Ibero-Romance*. Amsterdam: John Benjamins, 153-180.

[6] **Feldhausen**, I. **& Vanrell**, M. 2015. Oraciones hendidas y otras estrategias de marcaje del foco en español. *Revista Internacional de Lingüística Iberoamericana* 13(2): 39-60.

[7] **Féry**, C. 2011. German sentence accents and embedded prosodic phrases. *Lingua* 121(13): 1906-22.

[8] **Frascarelli**, M. 2000. *The Syntax-Phonology Interface in Focus and Topic Constructions in Italian*. Dordrecht: Kluwer.

[9] **Frota**, S. **& Prieto**, P. 2015. *Intonation in Romance*. Oxford: OUP.

[10] **López**, L. 2009. *A Derivational Syntax for Information Structure*. Oxford: OUP.

[11] **Pešková**, A. 2015. *Sujetos pronominales en el español porteño*. Berlin: De Gruyter.

[12] **Vanrell**, M. **& Fernández-Soriano**, O. 2013. Variation at the Interfaces in Ibero-Romance. *CatJL* 12, 253-282.

[13] **Zubizarreta**, ML. 1999. Las funciones informativas: Tema y foco. In *GDLE,* I. Bosque & V. Demonte (eds), pp. 4215–4244.

Monday, 16:00 - 17:00
Poster Session 2, #11

Phonatory and duration effects of focus position on the right phrase edge in South Connaught Irish

Maria O'Reilly, Ailbhe Ní Chasaide
Phonetics and Speech Laboratory, School of Linguistic, Speech and Communication Sciences, Trinity College Dublin, Ireland

Introduction: The work on the intonation of focus in Irish has shown that the same pitch accent type is used on the nuclear accent in broad and contrastive focus alike [1-3]. Also observed were *f0* adjustments in the IP, i.e. boosting of the focal peak, post-focal de-accentuation and reduction of the post-focal *f0* range [1-3], similarly to e.g. [4]. Recently, it has been reported for Ulster Irish that the IP-final syllable undergoes devoicing when focus occurs early in the phrase, especially in the IP-initial position [5].

This study examines the phonatory and durational effects of focus location on the IP-final syllable in South Connaught Irish, and particularly the occurrence of devoicing and creaky voicing in the IP-final syllable, often associated with low *f0* [6] as occurs at the right IP edge. If, as in Ulster Irish, devoicing/creaky voicing is more extensive with early focus, this may indicate that the voice source is weakening over the postfocal material, as a potential correlate of postfocal deaccentuation. If, on the other hand the effects are not more extensive with earlier focus, one would infer that these phonatory shifts are simple boundary effects.

Materials & methods: The corpus was recorded for four native speakers of South Connaught Irish from Cois Fharraige in County Galway in the west of Ireland. The data in this study include broad focus and contrastive focus renderings of a simple declarative shown in (1):

		A1		**A2**		**A3**
(1)	Bhí	**Méabh**	ina	**luí**	ar an	**lea**ba.
	/vʲi	ˈmʲev	nˠə	ˈlˠi	ərˠ ə	ˈʎabˠə/
	Was	*Méabh*	*in her*	*lying*	*on the*	*bed.*

'Méabh was lying on the bed.'

Broad (bf) and contrastive focus on one of the three accentable items (fA1, fA2, or fA3) were prompted in a question-answer context. A total of 64 tokens were obtained, out of which the fA2 data of Speaker 1 were excluded due to failure to produce focus in this context. The data were analysed for phonation type (devoicing and creak) in the IP-final syllable by means of auditory and visual (spectrogram) analysis. Furthermore, intensity and duration of the IP-final vowel, as well as speech rate in the post-focal material were measured.

***Results*:** The auditory analysis suggests that phonation type in the IP-final vowel is related to focus position in the IP. Specifically, the vowel is more likely to be devoiced the earlier focus occurs in the IP (Figure 1). Interestingly, phonation strategies appear speaker-specific: only Speakers 1 and 4 devoice, Speaker 3 exhibits infrequent instances of creaky voice across all focus conditions, while Speaker 2 does not resort to devoicing *or* creak whatsoever.

The acoustic analysis (Figure 2a-c) reveals that (a) intensity of the IP-final vowel is lower when focus occurs early in the IP, i.e. in fA2, and especially fA1; (b) duration of the final vowel is not affected by focus position, and (c) speech rate in the post-focus material is higher when focus occurs early in the IP.

Conclusions: In South Connaught Irish, the incidence of devoicing appears governed by the location of focus (cf. [5]), although it is not consistent with all speakers. For these speakers, it seems likely to indicate phonatory correlates of postfocal deaccentuation. Creaky voicing, when it occurs, is not limited to a specific focus type or position, and appears to be a simple phrase-edge effect. Intensity and speech rate are correlated with focus position (cf. [7]). Together, these combine with *f0* in the realisation of postfocal deaccentuation, contributing thus to the signalling of focus in the sentence.

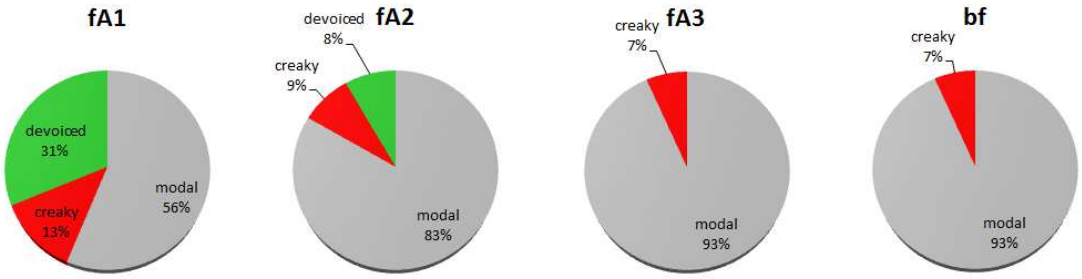


Figure 1. *Frequency distribution of phonation types (modal, creaky and devoiced) in the focus data for South Connaught Irish. fA2 includes data from three speakers.*

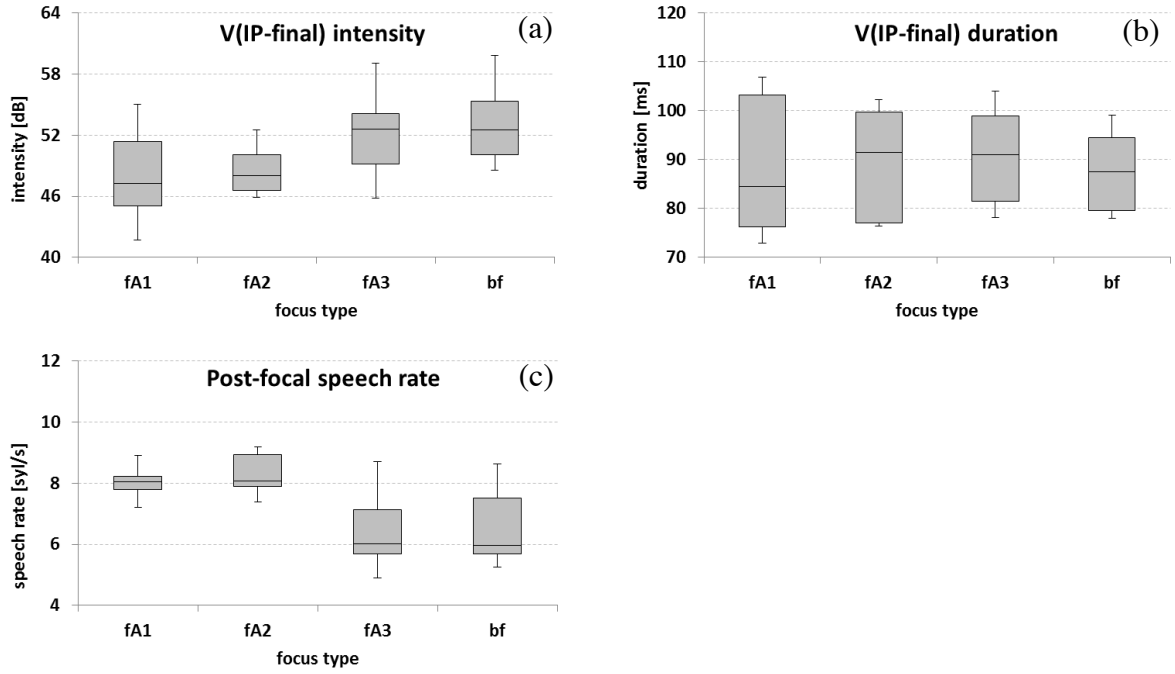


Figure 2. *Box plots for (a) intensity of the IP-final vowel, (b) duration of the IP-final vowel, and (c) post-focal speech rate in the focus data for South Connaught Irish.*

References

[1] O'Reilly, M., Dorn, A., & Ní Chasaide, A. 2010. Focus in Donegal Irish (Gaelic) and Donegal English bilinguals. Proceedings of *Speech Prosody 2010*, Chicago, USA.

[2] Dorn, A., & Ní Chasaide, A. 2011. Effects of focus on *f0* and duration in Donegal Irish declaratives. Proceedings of *Interspeech 2011*, Florence, Italy.

[3] O'Reilly, M., & Ní Chasaide, A. 2016. The prosodic effects of focus in the Irish of Cois Fharraige. Paper presented at *TIE 2016*, Canterbury, UK.

[4] Xu, Y., & Xu, C. 2005. Phonetic realization of focus in English declarative intonation. *Journal of Phonetics*, 33, 159-197.

[5] Dorn, A., & Ní Chasaide, A. 2016. Devoicing in post-focal constituents in Donegal Irish. Paper presented at *TIE 2016*, Canterbury, UK.

[6] Lindblom, B. 2009. F0 lowering, creaky voice, and glottal stop: Jan Gauffin's account of how the larynx works in speech. Proceedings of *FONETIK 2009*, Stockholm, Sweden.

[7] Nadeu, M., & Vanrell, M. M. 2015. Postfocal material in sentences with contrastive focus in Catalan and Spanish. Paper presented at *PAPE 2015*, Cambridge, UK.

/r/ is for rough: Sound symbolism for English touch concepts

Bodo Winter[1], Márton Sóskuthy[2] & Marcus Perlman[3]
[1]*University of Birmingham,* [2]*University of York,*
[3]*Max Planck Institute for Psycholinguistics, Nijmegen*

The sound structure of spoken words often expresses sensory imagery through iconicity, including onomatopoeic and cross-modal correspondences between sound and meaning [1-2]. This study examines the association between the phoneme /r/ and tactile meanings related to roughness in English and Hungarian, a non-Indo-European language. We propose that there is a perceived cross-modal mapping between the irregularity of rough surfaces and the repeated interruption of airflow involved in producing trilled /r/. Notably, the phonetic realization of /r/ was a trill for most of the history of English [3, p. 229], and /r/ is often realized as a trill in Hungarian.

To measure the roughness-based iconicity of words in English, we used native speaker ratings of roughness collected by [4] for sensory adjectives that describe surface properties, e.g., *rough, prickly and smooth*. For example, the adjective *barbed* has a high roughness rating (+6.3) compared to *lubricated* (–6.2). We phonemically transcribed the 124 normed adjectives from this list and calculated form-based distance (e.g., *satiny* and *sandy* are phonologically closer than *jagged* and *oily*). We correlated these distances with distances in semantic space (defined by roughness), using Mantel tests (cf. [5]). Semantic distances in English were significantly correlated with form-based distances, indicating that words similar in roughness share similar sounds. The analysis was replicated for Hungarian surface descriptors, based on translations of the norms [4] for English.

A deeper analysis of the correlation structure shows that the phoneme /r/ contributes disproportionately to the correlation between form-based distances and meaning distances in both languages. In English, 65% of predominantly rough words contain /r/, compared to only 34% of smooth words. In a follow-up experimental study, we showed that 60 native speakers of American English generalized this pattern to novel words such as *rorce* and *snilm*, where /r/ was associated with rough meanings. Together our results provide evidence that the "/r/ for rough" pattern is present in both languages and also that it is used productively by speakers of English. The fact that these two unrelated languages exhibit the same pattern suggests that it is rooted in iconicity.

[1] Dingemanse, M. 2012. Advances in the cross-linguistic study of ideophones. *Language and Linguistics Compass*, 6, 654–672.

[2] Cuskley, C., & Kirby, S. 2013. Synaesthesia, cross-modality and language evolution. In Simner, J. & Hubbard E.M. (Eds.), *Oxford Handbook of Synaesthesia*. Oxford: Oxford University Press, 869–907.

[3] Prins, A. A. 1972. *A History of English Phonemes: From Indo-European to Present-Day English*. Leiden: Leiden University Press.

[4] Stadtlander, L. M., & Murdoch, L. D. 2000. Frequency of occurrence and rankings for touch-related adjectives. *Behavior Research Methods, Instruments, & Computers*, 32, 579–587.

[5] Monaghan, P., Shillcock, R. C., Christiansen, M. H., & Kirby, S. 2014. How arbitrary is English? *Philosophical Transactions of the Royal Society of London: Series B, Biological Sciences*, 369, 20130299.

Monday, 16:00 - 17:00
Poster Session 2, #13

After Low High…Low High? A study on basic prosodic unit right boundaries in Korean

Inyoung Kim and Maelle Amand
EA 3967, CLILLAC-ARP, Université Paris Diderot

Korean Prosody has an Accentual Phrase level with an LHLH underlying tone and requires at least four syllables to realize all four tones (Jun 2000). Seoul Korean has no lexical accent, and the Seoul dialect doesn't have pitch accent. In this study on Korean read speech, we analyze the realizations of the second L and H tones in detail verifying the segmental anchoring hypothesis (Ladd *et al*. 1999, Arvaniti *et al*. 1998).

In a study on read speech, Kim (2015) showed that the first L and H tones have an anchoring effect: the first L tone was realized on the first syllable with a constant distance, around 30ms after the beginning of the first segment, and the first H tone was realized on the second syllable with a constant distance around 20ms from the onset of the vowel. The second L and H tones are realized somewhere on the last two syllables. Cho and Flemming (2015) showed that the realization of the second L tone under time pressure is analyzed to be "undershot" and this is considered as a compression process. Regarding Accentual Phrase of Seoul Korean within the same theoretical framework on prosody, French shares same aspects despite typological distance. French which has no lexical accent is also reported to have Accentual Phrase with underlying LHLH tones (Jun and Fougeron 2002). The first two L and H tones and the second L and H tones are generally realized as F0 rising, where the first LH is analyzed as bitonal phrase accent and the last LH as bitonal pitch accent in Welby (2006). Her study shows that the first L is realized near the lexical word onset, but the second L of the final rising didn't show any consistent anchoring effect. Impressionistic analysis of Korean prosody containing four syllables might show two rises, but the nature of Korean Accentual Phrase final LH is yet to be clear.

Our study investigates realizations of the second L and H tones in Korean. The corpus consists 175 target lexical words which are realized as four syllable Accentual Phrase. These targets are at the initial position in two different carrier phrases in order not to repeat the same phrase (350 tokens per speaker). Each sentence was presented visually and the speakers were asked to read twice, namely, at a normal and a fast speed. Participants had a 2 to 3 minute rest after every 100 phrase. All the target sentences were presented randomly with 136 filler phrases. Four female and male speakers ranging from the age of 22 to 29 and 22 to 33 respectively, participated in the experiment. The recording was carried out in a recording room. Regarding the data analysis, we measured the segmental maximum and minimum F0 points and within the final syllable where the second pair of L and H tones are expected to be realized. At normal speed, the minimum F0 was generally realized at the boundary between the last two syllables. For a more detailed analysis, we selected the two segmental points for comparisons. A non-parametric paired test revealed a significant difference of F0 occurring between the offset of the penultimate vowel and the onset of the last vowel (fig. 1). The segmental point of maximum F0 also revealed a significant difference, especially in faster speech where we also observed a general Accentual Phrase F0 contour difference (fig. 2).

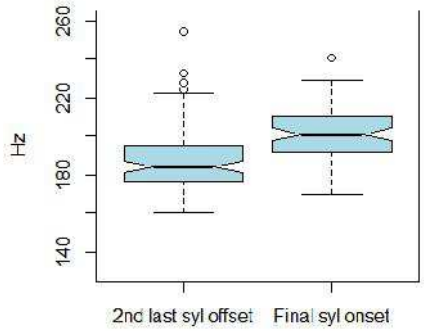


Figure 1. *Comparison between the offset of the penultimate syllable and the onset of the last syllable by a female speaker.*

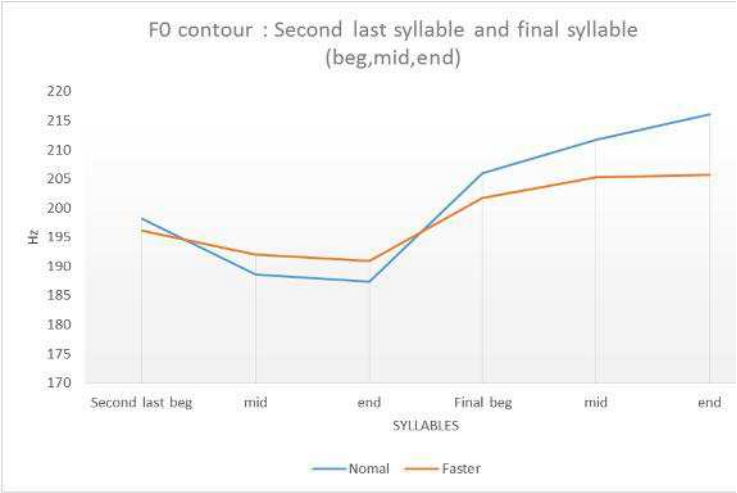


Figure 2. *F0 contour on the penultimate and last syllables by a female speaker, normal vs. faster speech.*

[1] Arvaniti, A., Ladd, D. R. & Mennen, I. 1998. Stability of tonal alignment: the case Greek prenuclear accents. *Journal of Phonetics*, 26.

[2] Cho, H. & Flemming, E. 2015. Compression and truncation: The case of Seoul Korean accentual phrase. *Studies in Phonetics, Phonology, and Morphology*, 21(2).

[3] Jun, S.-A. 2000. K-ToBI (Korean ToBI) Labelling Conventions. http://www.linguistics.ucla.edu/people/jun/ktobi/k-tobi.html

[4] Jun, S.-A. & Fougeron, C. 2002. Realizations of accentual phrase in French intonation. *Probus*, 14(1).

[5] Kim, S. 2015 Effects of syllable structure and prominence on the alignment and the scaling of the phrase-initial rising tone in Seoul Korean: A preliminary study. *Journal of the Korean Society of Speech Sciences*, 7(4).

[6] Ladd, D. R., Faulkner, D., Faulkner, H. & Schepman, A. 1999. Constant "segmental anchoring" of F0 movements under changes in speech rate. *Journal of the Acoustical Society of America*, 106(3).

[7] Welby, P. 2006. French intonation structure: Evidence from tonal alignment, *Journal of Phonetics*, 34(3).

The prosody of different question strategies in Labourdin Basque and the syntax-phonology interface

Aritz Irurtzun & Maia Duguine
CNRS-IKER & UPV/EHU

1. Introduction: Basque is a head-last and discourse-configurational language that is typically taken to be [+wh movement]. However, [1] showed that young speakers of the Navarro-Labourdin dialect have developed a *wh in situ* strategy with a range of different syntactic and semantic properties (see (3)). This paper provides an empirical assessment of the prosody of the different interrogative strategies of Navarro-Labourdin Basque (a virtually unstudied stress-accent variety spoken in France) with an experiment contrasting the intonation of wh-movement, wh in situ and polar questions to test whether the claim by [2] that wh in situ (in French) is licensed by imposing the polar question tune to wh in situ constructions can also be extended to Basque (a language in a diglossic situation wrt French).
2. The experiment: The questionnaire was composed of three different conditions (wh in situ, wh-movement and polar question constructions) as (2)-(4) and 9 items in total. 6 female participants were asked to produce these sentences in as natural way as possible. Each participant produced 3 renditions of each item, which amounts to a total of 162 utterances (6 speakers x 9 items x 3 renditions).
3. Measurements: We measured duration, and intensity and F0 maxima, minima and means in each of the accented syllables (subject, object, verb) and F0 maxima of the last two syllables of each utterance in order to assess the range of final rise. This amounts to a total of 9 measurements per utterance.
4. Results and discussion: Large differences were observed across conditions (see time-normalized pitch contours for wh-movement (5), polar questions (6), and wh in situ (7) questions). Acoustic measurements in the three accented syllables (subject, object, verb) showed a range of statistical differences in F0 valuess, but not in intensity nor duration: at the subject (s1), F0 maxima values are higher in the wh in situ condition (μ=271.89; σ=39.54) than in the wh-movement (μ=256.68; σ=36.75) and the polar question (μ=255.94; σ=43.68) conditions. Continuing only with the in situ and polar questions (the movement questions have a different word order, see (2)-(4)), we see that at the object (s2) the situation reverses and F0 maxima are significantly lower in the wh in situ condition (μ=231.56; σ=13.90) than in the polarQ (μ=248.03; σ=28.34). Last, at the lexical verb (s3), the polar question condition has much higher f0 maxima values (μ=237.34; σ=32.18) than the wh in situ condition (μ=206.70; σ=10.35), see (8). So we could think that since the tunes of in situ questions and polar questions are different, we could not extend [2]'s analysis to Navarro-Labourdin Basque. However, the study of F0 movements in the last two syllables will show a different picture, for polar questions and wh in situ show a very similar rise (in semitones): wh in situ (μ=7.50; σ=3.50); polarQ (μ=7.32; σ=2.06), very different from the values of wh-movement questions (see (9)). A Bayesian MCMC model with 100,000 possible parameters confirmed the difference between the polar question condition and wh-movement condition as well as the difference between the wh in situ and the wh-movement condition. But, interestingly, it did not reveal a statistically credible difference between the polar question condition and the wh in situ condition. Mean values being similar, the Bayesian estimation of differences did not allow either to declare that final rises in wh in situ and polar question conditions are equal, but a substantial part of the highest density interval for mean values in the model fell inside of the "region of practical equivalence" (see [3]), which leads us to think that these are very similar rises. We will argue that the differences observed in the accents are not very different from those observed in French in the region preceding the final rise (see [4],[5]), and therefore, that [2]'s suggestion could be naturally applied to Basque too.

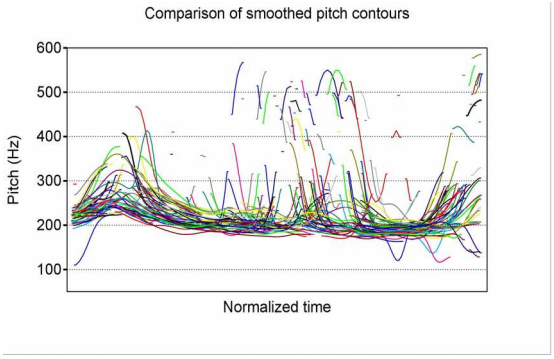
(1) Nik liliak eraman ditut.
I.erg flowers bring aux
Jon brought the flowers.

(2) Nok eraman ditu liliak? **[movement]**
who.erg bring aux flowers
"Who took the flowers?"

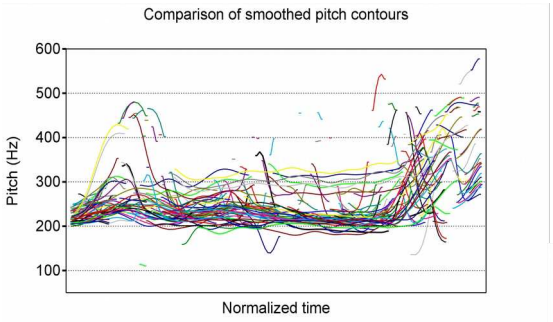
(3) Nok liliak eraman ditu? **[in situ]**
who.erg flowers bring aux
"Who took the flowers?"

(4) Nik liliak eraman ditut? **[polar Q]**
I.erg flowers bring aux
"Did I bring the flowers?"

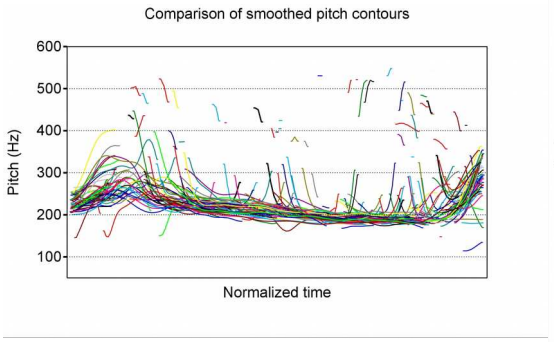
(5)



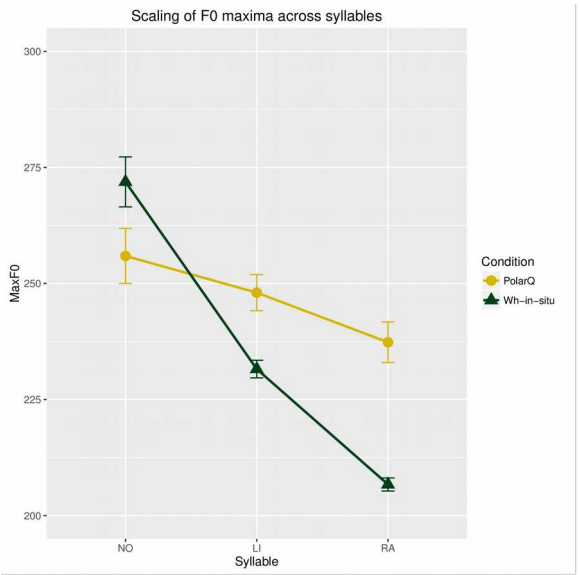
(6)



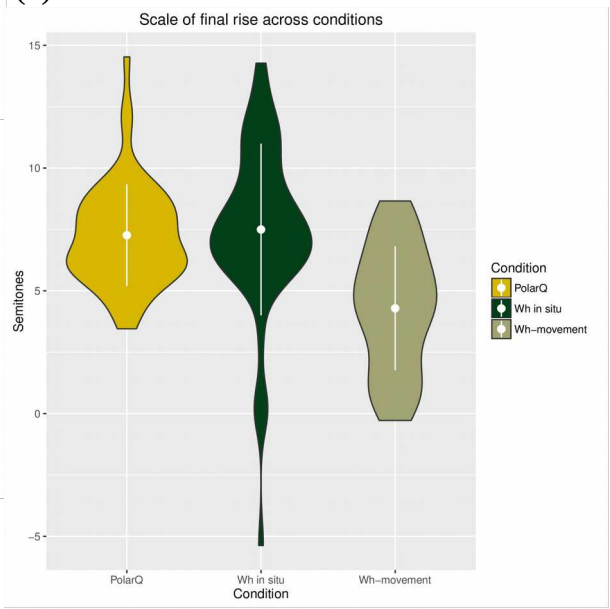
(7)



(8)



(9)



[1] Duguine, M., & A. Irurtzun. 2014. From obligatory Wh-movement to optional Wh in situ in Labourdin Basque. DOI: 10.1353/lan.2014.0006

[2] Cheng, L. & J. Rooryck. 2000. Licensing wh-in-situ. DOI: 10.1111/1467-9612.00022

[3] Kruschke, J. 2015. Doing Bayesian Data Analysis. London: Academic Press.

[4] Déprez, V., K. Syrett & S. Kawahara. 2013. The interaction of syntax, prosody, and discourse in licensing French wh-in-situ questions. DOI: 10.1016/j.lingua.2012.03.002.

[5] Gryllia, S., L. Cheng & J. Doetjes. 2016. On the intonation of French wh-in-situ questions: What happens before the wh-word is reached? DOI: 10.21437/SpeechProsody.

The effect of ultrasound and video feedback on the production and perception of English liquids by French learners

Hannah King & Emmanuel Ferragne
Clillac-ARP, Université Paris Diderot

Most varieties of English distinguish two allophones for /l/: a clear variant in onset, and a dark one in coda. French, however, has one clear allophone in all contexts [1]. Articulatory data has shown that English coda /l/ is produced with a retraction and lowering of the tongue body [2]. English /r/ is generally divided into two broad categories, retroflex [ɻ] or bunched [ɹ], which, unlike /l/ allophones, have been shown not to make a perceptible difference to native listeners [3]. /r/ exhibits therefore a "many-to-one articulatory-acoustic relationship" [4]. The French rhotic consonant, the uvular fricative [ʁ], is dissimilar in both acoustic and articulatory terms. Previous research [5] has revealed that French learners struggle to differentiate English /r/ from /w/. A purely phonological modelling of the perception of non-native contrasts cannot account for such a finding. Language-specific phonetic similarities may then also play a role. We predicted that at an articulatory-phonetic level, French learners would not produce English liquids in the same way as native speakers. To help French learners acquire the complex articulatory gestures for these sounds, we propose an articulatory training session using ultrasound tongue imaging (UTI) and video of the lips as visual biofeedback. If the source of perceptual constraints lies in surface-level articulatory phonetics, an improvement in production should result in an improvement in perception of these contrasts.

An experimental group (EG) of seven learners had 30 minutes of individual pronunciation training with biofeedback using an English native speaker's productions as a model. A control group (CG) of seven learners underwent a repetition task without biofeedback. The learners' articulatory and acoustic data were recorded before and after training. AB discrimination tasks were also carried out. Learners heard word pairs contrasting [ɹ]-[w] and [l]-[ɬ] produced by two native speakers. 15 native speakers judged the learners' productions before and after training. Ultrasound data was analysed by visually inspecting the tongue contours. Lip videos were assessed for rounding and formant values from F1-F3 were extracted.

None of the learners used tongue retraction or lowering for /l/ in coda prior to training. After training, however, all participants in EG and two in CG did so. This correlated with a significantly lower F2 ($p<0.0001$) and higher native speaker ratings ($p<0.0001$). Interrater conformity was deemed good (ICC=0.632). Nearly all learners produced either a retroflex (10/14) or a bunched (2/14) /r/ prior to training with some lip rounding (13/14). One learner used both configurations. Another learner used [ʁ] prior to training with biofeedback but produced [ɻ] after training. No significant acoustic difference was observed between [ɻ] and [ɹ] but both had a significantly lower F3 than [ʁ] ($p<0.001$). Native judges significantly preferred [ɻ] or [ɹ] to [ʁ] ($p<0.0001$) with no significant preference for either [ɻ] or [ɹ]. As all learners who followed training with biofeedback acquired more native-like productions for coda /l/ and /r/ (if previously [ʁ]), we conclude that training with biofeedback improved production.

No significant difference was found in the learners' discrimination of [ɹ]-[w] and [l]-[ɬ] before and after training or between the two groups. We conclude that an improvement in production does not necessarily result in an improvement in perception. It seems unlikely then that L2 speech perception is purely phonetically motivated. However, we argue that phonetics does indeed play a role. Unexpectedly, the learners who used [ɹ] before training could distinguish [ɹ] from [w] better than those who used [ɻ] ($p<0.01$). Given their relatively similar articulatory gestures (figure 1), it seems that [ɹ] is intrinsically more [w]-like than [ɻ]. We speculate that the perceptual system of those who bunch can distinguish finer phonetic detail to those who retroflex. This conclusion is additionally supported by the fact that the learner who used both retroflex and bunched tongue configurations had on average the highest d prime score for discriminating /r/ from /w/. More data is needed to further verify this claim.

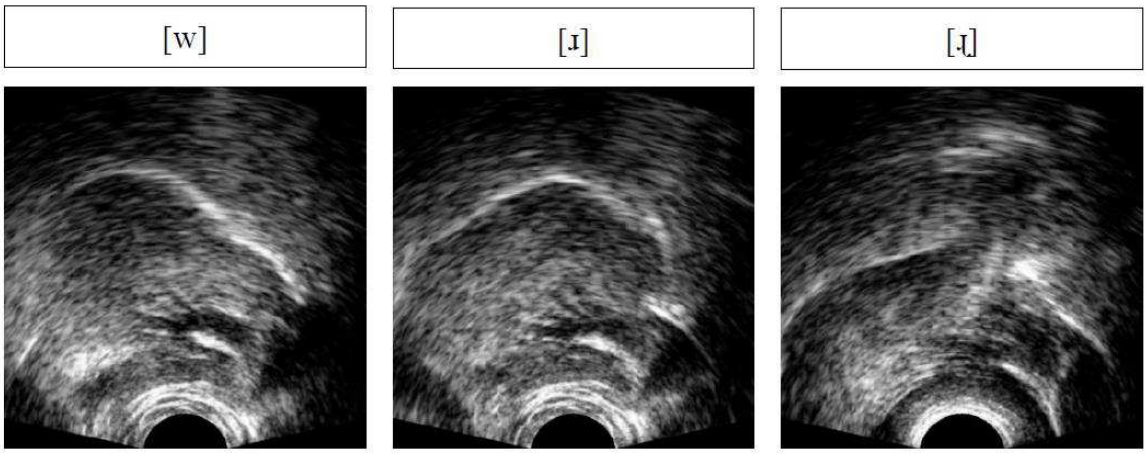


Figure 1. *Midsagittal ultrasound images from native speakers (posterior on the left)*

[1] Mielke, J., Baker, A. & Archangeli, D., 2016. Individual-level contact limits phonological complexity: Evidence from bunched and retroflex /ɹ/. *Language*, 92(1), pp.101-140.

[2] Sproat, R. & Fujimura, O., 1993. Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of phonetics*, 21(3), pp.291-311.

[3] Delattre, P. & Freeman, D.C., 1968. A dialect study of American r's by x-ray motion picture. *Linguistics*, 6(44), pp.29-68.

[4] Espy-Wilson, C. Y., Boyce, S. E., Jackson, M., Narayanan, S., & Alwan, A. (2000). Acoustic modeling of American English /r/. *The Journal of the Acoustical Society of America*, 108(1), pp.343-356.

[5] Hallé, P.A., Best, C.T. & Levitt, A., 1999. Phonetic vs. phonological influences on French listeners' perception of American English approximants. *Journal of Phonetics*, 27(3), pp.281-306.

Monday, 17:30 - 18:00
Oral Session 3, #2

L2 status affects L3 learning for the onset of acquisition:
A developmental study of L1 English, L2 Spanish, and L3 Catalan

Nicholas Henriksen & Sarah K. Harper
University of Michigan & University of Southern California

Research in third language (L3) phonological acquisition has focused on how a learner's L3 interacts with previously acquired linguistic systems (L1 and L2) and how these interactions vary as a consequence of factors like the degree of phonetic similarity between target phones in each language, speakers' L2 proficiency, and the typological similarity between the L1, L2 and L3 (see review in [1]). In particular, the findings of this research show strong support for the *L2 status effect* [2], whereby the learner's L2 is preferentially selected as the source of influence into the L3 under all circumstances. However, it is unclear whether L2 status effects are present at the **initial stage** of L3 learning or instead emerge as the learner gains greater proficiency in and metalinguistic awareness of the L3 system. It is also unclear whether the **phonetic similarity** of the phone in question (i.e., comparing target L1, L2, and L3 production) plays a role in activating or inhibiting L2 status effects.

The goal of this longitudinal study is to understand the production of L3 phones that differ in their phonetic similarity to learners' L1 and the L2. Specifically, we examine the production of intervocalic /b/ and word-final /l/ by L1 English/L2 Spanish speakers during their first semester of L3 Catalan, comparing the realization of these phones in their L1, L2, and L3 at the beginning and the end of their term of Catalan coursework. These sounds were selected because of their relationship to the learners' L1 and the L2 in terms of phonetic realization and phonological patterning. On the one hand, Spanish (L2) and Catalan (L3) have an allophonic pattern of intervocalic voiced stop lenition (/b d g/ > [β ð ɣ]) that does not occur in English (L1). On the other hand, English (L1) and Catalan (L3) share a pattern of coda /l/ velarization that Spanish (L2) lacks.

We recruited 20 university-level L1 English-speaking L2 learners of Spanish. Eleven participants (the experimental group) were L2 Spanish/L3 Catalan learners enrolled in their first semester of Catalan study. The remaining nine learners (control group) were L2 Spanish learners with no L3 exposure. All learners read aloud carrier sentences in their L1, L2, and (for the experimental group) L3 at the beginning and end of an academic semester. The carrier phrases were blocked by language, and the order of presentation counterbalanced between recording sessions. The target words for each language included 48 instances of intervocalic /b/ (Spirantization Targets) and 48 instances of word-medial or -final /l/ (Velarization Targets). For /b/, we extracted the C/V intensity ratio comparing the minimum intensity of /b/ and the maximum intensity of the following vowel, commonly used in research on consonant lenition in Romance languages (e.g., [3]). For /l/, we extracted F1 and F2 values at midpoint to gauge the degree of velarization (e.g., [4]).

Our results show that L3 Catalan learners produced both /b/ [Fig. 1] and /l/ [Fig. 3] with greater phonetic similarity to their own productions of Spanish /b/ and /l/ than to English at the beginning and the end of the semester despite the presence of similar /l/ velarization patterns in L1 English and L3 Catalan. This mirrors patterns observed in L3 learners with a greater period of exposure to the language [1], and precludes the observation that there was no evidence for an effect of recording time on learners' production in any of the three languages [Figs. 2 & 4]. Additionally, we did not find significant differences between the experimental and control groups' production of /b/ and /l/ at either recording time, indicating that L3 learners' relatively short exposure to Catalan did not affect their L1 or L2 production. Altogether, our findings suggest that models of phonological acquisition [e.g., 5] may have to account for a lack of disparity in the behavior of "same" and "different" sounds for typologically similar L2-L3.

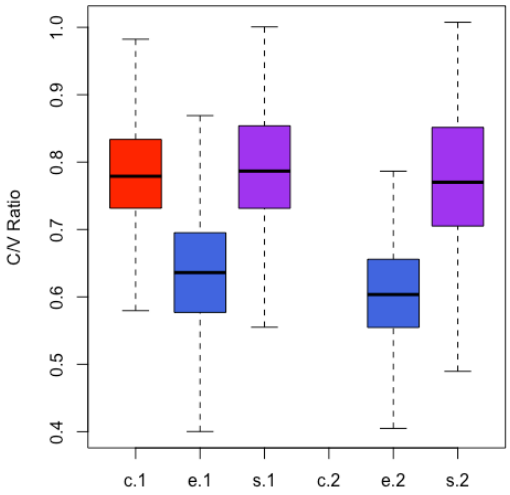


Figure 1. *Between-groups comparison of C/V Intensity Ratio values for lenition targets in all languages (Experimental Group on left (1), Control Group on right (2); Red = Catalan, Blue = English, Purple = Spanish)*

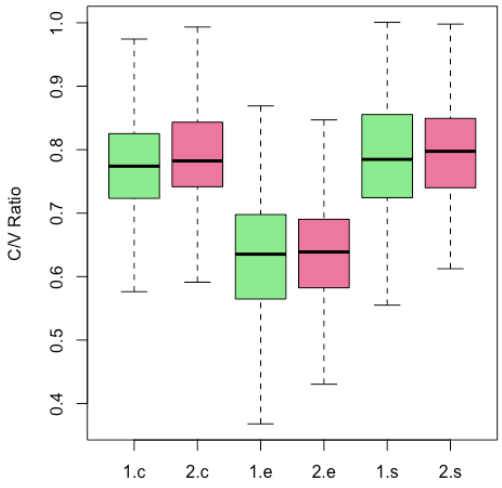


Figure 2. *C/V Intensity Ratio values for Experimental Group lenition targets by recording time and language (From L to R: Catalan, English, Spanish; Green = Time 1, Pink = Time 2)*

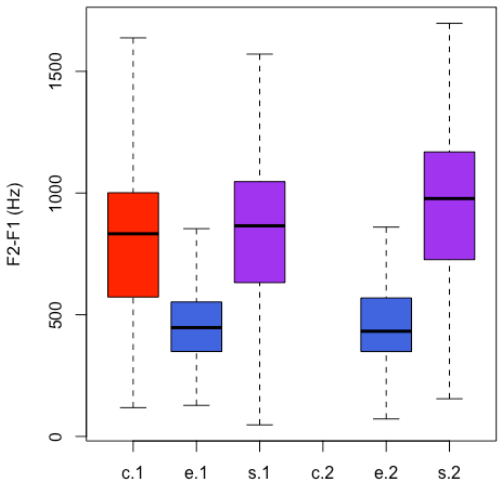


Figure 3. *Between-groups comparison of F2-F1 values for velarization targets in all languages (Experimental Group on left (1), Control group on right (2); Red = Catalan, Blue = English, Purple = Spanish)*

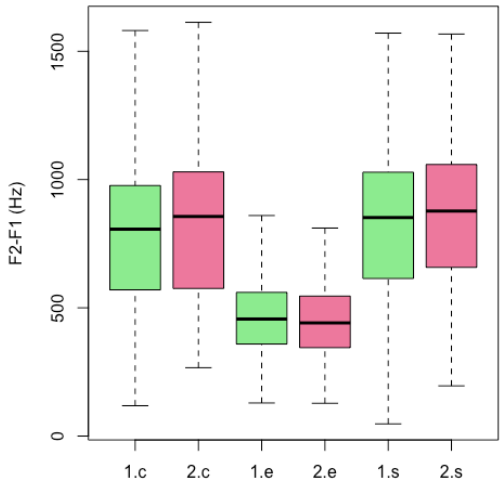


Figure 4. *F2-F1 values for Experimental Group velarization targets by recording time and language (From L to R: Catalan, English, Spanish; Green = Time 1, Pink = Time 2)*

[1] Cabrelli Amaro, J. (2012). L3 Phonology. In J. Cabrelli Amaro et al. (Eds.), *Third Language Acquisition in Adulthood* (pp. 33-60). Amsterdam: John Benjamins.
[2] De Angelis, G. (2007). *3rd or Additional Language Acquisition.* Clevedon: Multilingual Matters.
[3] Hualde, J. I., Simonet, M., and Nadeu, M. (2011). Consonant lenition and phonological recategorization. *Laboratory Phonology* 2(2), 301-329.
[4] Simonet, M. (2010). Dark and clear laterals in Catalan and Spanish: Interaction of phonetic categories in early bilinguals. *Journal of Phonetics* 38, 663-678.
[5] Flege, J. (1995). Second-language Speech Learning. In W. Strange (Ed). Speech Perception and Linguistic Experience (pp. 229-273). Timonium, MD: York Press.

Articulatory correlates of French and English metrical structure: Influences from L1

Donna Erickson[1], Caroline Smith[2], and Christophe Savariaux[3]
[1]Kanazawa Medical University, [2]University of New Mexico, [3]GIPSA-Lab

All spoken languages show rhythmic patterns, which can be described in terms of Metrical Phonology [1, 2]. Recent work in several languages [3, 4, 5] shows that metrically assigned stress levels of the utterance correlate significantly with amount of jaw displacement, with corresponding changes in F1 values. These patterns appear to be independent of pitch accent or tonal contrasts [4, 5]. As rhythm is a fundamental component of a language's phonology, changing to a different rhythm when learning a second language is likely to be a challenging task, but can be achieved as fluency increases [6, 7].

This paper asks whether the articulatory (jaw displacement) patterns of French speakers also match the metrical structure, and extends this idea to ask (a) how first language rhythm affects second language rhythm, and (b) how does the level of fluency in the second language affect language rhythm, in both first and second languages. We compare English speakers' productions of French and French speakers' productions of English, where the speakers have various levels of proficiency in their second language, ranging from bilingual to poor, ages from 22-72.We used the NDI Wave articulograph to record jaw displacement patterns of 3 English speakers (2 m) and 5 French speakers (3 m) reading English and French sentences. The diagram below is a possible metrical structure for one of the French sentences.

Intonational Phrase															x
Accentual Phrase			x				x		x		x				x
Syllable	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
IPA	na	ta	ʃa	na	ta	ʃa	pa	sɔ̃	ʃa	pa	ʃa	ki	se	ʃa	pa

Natacha n'attacha pas son chat Pacha qui s'échappa
'Natasha didn't attach her cat Pacha who escaped.'

The results suggest that (1) native speakers' jaw displacement patterns for the French sentence correspond closely to the metrical structure of the sentence, with a significant correlation between metrical structure and amount of jaw displacement/F1, and (2) there is an effect of first language prosody on second language prosody: Poor second language speakers tend to carry over their first language jaw displacement patterns. For E3 (a poor speaker of French), jaw displacements were largest on the syllables that would be stressed in English, i.e., second syllable of *Natacha*; this contrasts with French speakers with largest jaw displacement on the final syllable of this word, as would be expected given its prominence at the end of an Accentual Phrase. More fluent second language speakers produced jaw displacement patterns that matched the native language speakers more closely; speaker E1's production of the French sentence shows larger jaw displacements at a subset of the locations where French speakers also made larger displacements, but not at locations predicted by English stress patterns. These patterns are illustrated by the bargraphs in Fig. 1 which show jaw displacement in mm for each syllable as measured from the occlusal biteplane for the above sentence for a native speaker of French, E1 and E3. Notice that even for the vowels [a], the jaw displacements varies; the native French speaker showed increased jaw displacement on syllables 3, 7, 9 and 15, which match all but one of the phrasal stresses for this sentence as shown above. A second French speaker showed an identical pattern, while a third speaker showed a similar pattern to speaker E1, shown in Figure 1.

These results suggest that more than one metrical pattern is possible for this sentence, even for native speakers. E3 showed increased jaw displacement on syllables 2, 5, 7, 9, and 14, suggesting that he got the phrasal stress on syllables 7 and 9, but not the others. Instead, he tended to put stress on phrase-penultimate syllables 2, 5, and 14, which is reminiscent of the English lexical stress patterns.

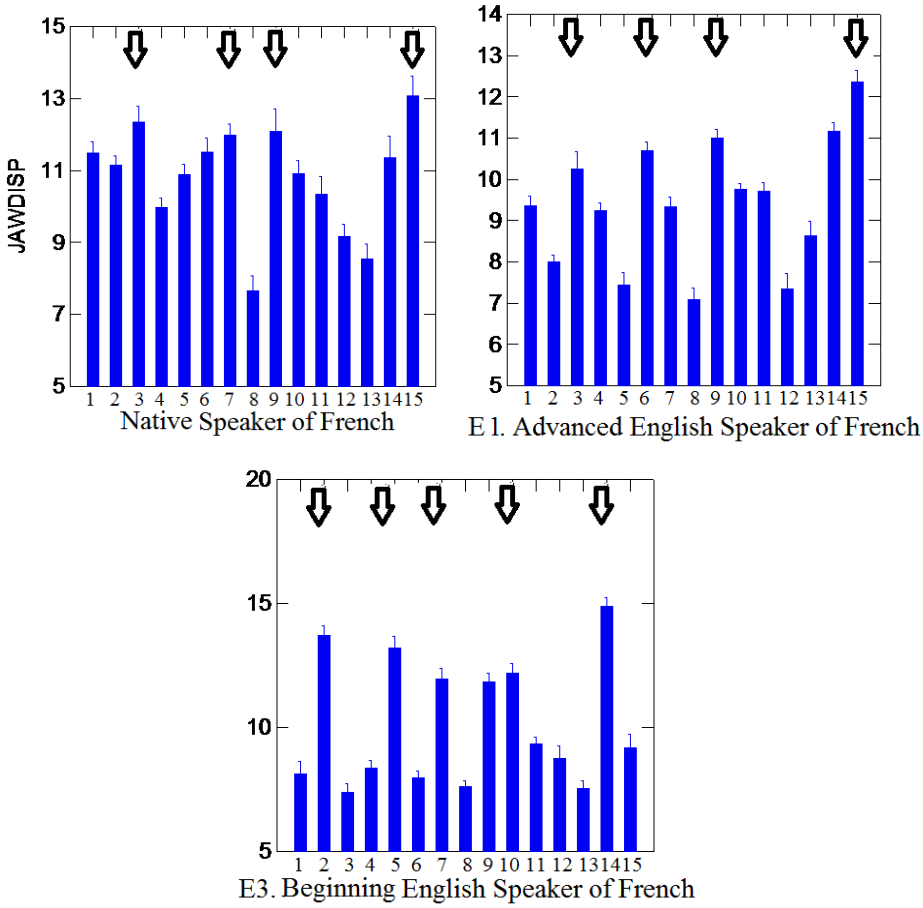


Figure 1. *Jaw displacement (mm) for each syllable in the sentence* Natacha n'attacha pas son chat Pacha qui s'échappa*. The error bars show standard error of the mean. Ordinate scaling is individual by speaker in order to better show the patterns. X-axis shows syllable numbers. Arrows indicate stressed syllables.*

[1] Liberman, M. and Prince, A. 1977. On stress and linguistic rhythm. *Linguistic Inquiry,* 8, 249–336.

[2] Hayes, B. 1995. *Metrical Stress Theory*. Chicago: University of Chicago Press,.

[3] Erickson, D., Suemitsu, A., Shibuya, Y., & Tiede, M. 2012. Metrical structure and production of English rhythm. *Phonetica,* 69, 180–190.

[4] Kawahara, S., Erickson, D., Moore, J., Shibuya, Y., & Suemitsu, A. 2014. Jaw displacement and metrical structure in Japanese: The effect of pitch accent, foot structure, and phrasal stress. *Journal of the Phonetic Society of Japan*, 18 (2), 77–87.

[5] Iwata, R., Erickson, D., Shibuya, Y., & Suemitsu, A. 2015. Articulation of phrasal stress in Mandarin Chinese. *Proc. of the Acoustical Society of Japan*, autumn meeting.

[6] Abe, Y., Wilson, I., and Erickson, D. 2012. Video recordings of L1 and L2 jaw movement: effect of syllable onset on jaw opening during syllable nucleus. *Journal of the Acoustical Society of America*, 132(3), (Kansas City meeting), 2005.

[7] Wilson, I., Erickson, D., & Horiguchi, N. 2012. Articulating rhythm in L1 and L2 English: Focus on jaw and F0. *Proc. of the Acoustical Society of Japan*, autumn, 319-322.

The effects of prosodic context on word segmentation in Korean

Hae-Sung Jeon[1] & Amalia Arvaniti[2]
[1]*University of Central Lancashire* [2]*University of Kent*

Word segmentation is an active process of continuous hypothesis-testing in search for the word in the mental lexicon which best matches the incoming acoustic information [1]. In this process, listeners are aided by a variety of cues, including prosodic cues [2] which can be classified into global rhythmic cues and local cues to prosodic boundaries [3], [4], [5]. The former allow listeners to infer boundaries by anticipation [6]; the latter help listeners interpret the organisation of the incoming signal by bringing their attention to changes in prosody [5]. The central question of the present study is whether such cues facilitate word segmentation in Seoul Korean (henceforth Korean). Korean is of interest because it does not have lexical stress; rather, its basic unit of prosodic organization is the accentual phrase (AP) which shows evidence of acting similarly to the foot in terms of rhythm ([7]; see [8] for a review). Based on these properties of Korean prosody, it was expected that both rhythm regularity (operationalized as consistent syllable count across phrases) and local phrasing cues would aid segmentation. These hypotheses were tested by means of a word-spotting experiment.

Seventy two native speakers of Korean (38 F, 34 M, aged 19 to 29) took part. They listened to strings of syllables and as soon as they heard a target (a disyllabic or trisyllabic Korean noun) they had to press a button and say the word out loud. The syllable strings were organized into four or five phrases (APs or IPs), each with a LH F0 pattern. SYLLABLE COUNT was either *regular*, with three trisyllabic phrases before the target, or *irregular*, with two phrases of four and five syllables before the target; the target was always in a trisyllabic phrase, giving a 3-3-3-3 vs. 4-5-3 structure for regular and irregular SYLLABLE COUNT respectively. LENGTHENING was (i) absent from all phrases (*no lengthening*), (ii) found at the end of all phrases (*pre-boundary lengthening*); (iii) present only before the phrase containing the target (*pre-target lengthening*). Thus, depending on the combination of SYLLABLE COUNT and LENGTHENING participants had at their disposal only rhythmic or only phrasal cues, or a combination of the two. Participants were tested individually in the Hanyang Phonetics & Psycholinguistics Laboratory, Seoul, after being randomly allocated into six experimental lists. Each list started with 12 practice items, followed by 48 test strings mixed with 48 fillers. The experiment ran on PsychoPy2 ver.1.83.01. Response accuracy and reaction times (for correct responses only) were analyzed in R [9], using mixed-effects logistic regression models and linear mixed-effects models respectively; participants and items were treated as random factors; SYLLABLE COUNT, LENGTHENING, and TARGET TYPE (disyllabic, trisyllabic) were the fixed factors.

Overall, trisyllabic targets were responded to significantly faster and more accurately than disyllables (see Figures 1 and 2). This can in part be attributed to the lower Phonological Neightborhood Density (PND) [10] of the trisyllabic targets, but could also be related to the fact that these targets were the same length as the phrases they were placed in. As a result, for trisyllabic targets, accuracy was at ceiling; however, final lengthening in all phrases did reduce reaction times. For disyllabic targets, irregular syllable count and lengthening before the target led to greater accuracy.

The results imply that both rhythmic consistency and lengthening play a role in word spotting and, by extension, in speech processing in Korean, as in other languages. However, the results also indicate the language specificity of these effects. First, the prosodic cues were used primarily with disyllabic targets, which were cognitively more demanding to process. Second, the facilitating effect of rhythmic consistency was weak, unlike in English (e.g. [6], [7], [11]), possibly because strict consistency is not present in spoken Korean phrases. Overall, rhythmic consistency had a facilitating effect when targets mapped onto phrases, confirming the importance of phrasal organization in Korean speech processing.

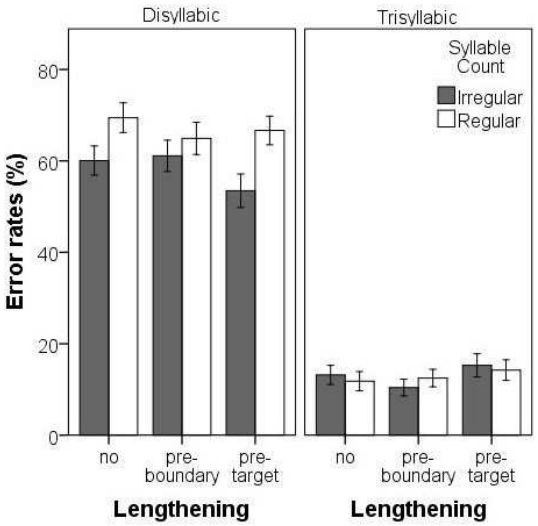


Figure 1. *Error rates (%). The error bars show ±1 standard error.*

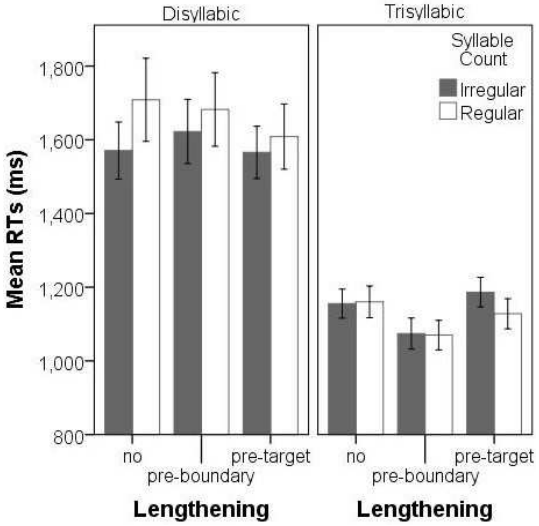


Figure 2: *Reaction Times (RTs) for correct responses. The error bars show ±1 standard error.*

References

[1] Cutler, A. 2012. *Native Listening: Language Experience and the Recognition of Spoken Words.* Cambridge, MA: MIP Press.

[2] Shattuck-Hufnagel, S., & Turk, A. E. 1996. A prosody tutorial for investigators of auditory sentence processing, *J. Psycholinguist. Res*. 25: 193–247.

[3] Fletcher, J. 2010. The prosody of speech: Timing and rhythm. In W.J. Hardcastle, J. Laver & F.E. Gibbon (Eds), *The Handbook of Phonetic Sciences*, pp. 523-602. Wiley-Blackwell: New York.

[4] Nolan, F., & Jeon, H.-S. 2014. Speech rhythm: a metaphor? *Phil. Trans. R. Soc. B*., 369. 20130396. http://dx.doi.org/10.1098/rstb.2013.0396

[5] White, L. 2014. Communicative function and prosodic form in speech timing, *Speech Commun*. 63-64: 38–54.

[6] Dilley, L. C., & McAuley, J. D. 2008. Distal prosodic context affects word segmentation and lexical processing, *J. Mem. Lang*. 59: 294–311.

[7] Tilsen, S. & Arvaniti, A. 2013. Speech rhythm analysis with decomposition of the amplitude envelope: Characterizing rhythmic patterns within and across languages, *J. Acoust. Soc. Am.* 134, 628–639.

[8] Jeon, H.-S. 2015. Prosody. In L. Brown & J. Yeon (eds), *Handbook of Korean Linguistics*, pp. 41-58. Wiley-Blackwell: New York.

[9] R Core Team. 2015. R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria. URL: http://www.R-project.org/

[10] Luce, P. A, & Pisoni, D. B. 1998. Recognizing spoken words: The neighbourhood activation model. *Ear. Hearing* 10, 1–36.

[11] Brown, M., Salverda, A. P., Dilley, L. C., & Tanenhaus, M. K. 2011. Expectations from preceding prosody influence segmentation in online sentence processing, *Psychon. Bull. Rev*. 18, 1189–1196.

Is the Cue of Pitch Rise to Spoken Word Segmentation Used in a Language-Specific or Cross-Linguistic Way? A Study of Listeners of Taiwanese Southern Min

Shu-Chen Ou and Zhe-Chen Guo
National Sun Yat-Sen University, Taiwan

The present study employs an artificial language learning (ALL) experiment to investigate the use of pitch rises in spoken word segmentation by native listeners of Taiwanese Southern Min (TSM). Listeners of all languages have to segment connected speech into its component words. Among the prosodic cues for achieving this, some have been reported to be cross-linguistic. For example, a rise in pitch is exploited by Korean (e.g., [1]) and French (e.g., [2]) listeners to locate the word-initial position and it has been argued to be a universal cue to boundaries ([3]). Yet, solutions to word segmentation problems may also be encouraged by language-specific lexical prosody; an example is that English listeners tend to treat stressed syllables as word beginnings ([4]). TSM stands out as a unique case for exploring the role of pitch rises in spoken word segmentation. The language has a robust tone sandhi process whereby the final syllable in the sandhi domain bears its citation tone whereas the other syllables take their corresponding sandhi tones, following the rules in (1).

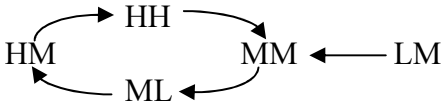
Given that the LM tone, the only rising tone in TSM, emerges only at the end of the sandhi domain and this position is typically the final position of a TSM word, it is hypothesized that TSM listeners would have developed a segmentation strategy that recognizes any pitch rises as cues to the word-final position. However, the cross-linguistic utility of pitch rises as observed in [1] and [2] provides support for assuming the opposite: pitch rises are word beginning cues. The competing hypotheses were tested with an ALL experiment.

The artificial language consisted of six "words," which were meaningless trisyllabic sequences (e.g., [pakime]). Recorded individually, their constituent syllables were manipulated so that they had a leveled F_0 contour and an equal duration. They were combined to construct the words for a "TP-only" condition, where the transitional probabilities (TPs) between syllables were the only cues to segmentation. Four additional "cued" conditions were created by superimposing a pitch fall or pitch rise on the initial or final syllables of half of the six words. They were referred to as "fall-initial," "fall-final," "rise-initial," and "rise-final" conditions. The former two were included to control for the possibility that any extra pitch cue in a particular position would be useful. The experiment was run by using E-prime 2.0. Subjects first heard 100 repetitions of each of the six words, which were concatenated pseudo-randomly without inserting any pauses, and figured out what the component words of the artificial language might be. They then received a two-alternative force-choice test, each trial of which presented a real artificial word and a nonword for them to determine which one was a word of the language. Twenty TSM speakers were assigned to each condition.

A mixed-effects logistic regression model revealed that as shown in Figure 1, only subjects of the rise-initial condition responded significantly more accurately than those in the TP-only one ($\beta = .24$, $SE(\beta) = .10$, $z = 2.29$, $p < .05$). As only half of the six words were given an extra prosodic cue in the four cued conditions, a model that took this factor into account and was fitted to only the responses in these conditions suggested that the cue facilitated segmentation in the rise-initial condition ($\beta = .49$, $SE(\beta) = .16$, $z = 3.03$, $p < .01$) but hindered it in the rise-final one ($\beta = -.57$, $SE(\beta) = .16$, $z = -3.63$, $p < .001$), as in Figure 2.

The findings support the hypothesis that TSM listeners segment the artificial language in a fashion that has been attested in several languages: pitch rises are a word-initial position cue. Their sensitivity to pitch cues at word beginnings may have been promoted by the tendency of post-boundary syllables in TSM, which can be word-initial, to show domain-initial strengthening effects such as F_0 range expansion ([5]). Further ALL research can be carried out to cross-linguistically examine whether pitch rises are a universal cue to word beginnings.

(1) Tone sandhi of Taiwanese Southern Min ([6]). The letters L, M, and H represent low, mid, and high pitch values, respectively.



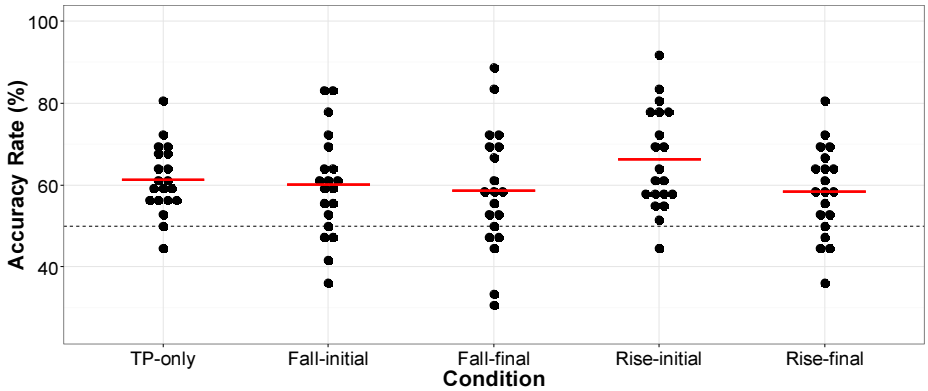


Figure 1. *Accuracy rate of each subject (indicated by a dot). The dotted line represents the chance performance level (i.e., 50% correct) and each red line segment the mean accuracy rate of each condition.*

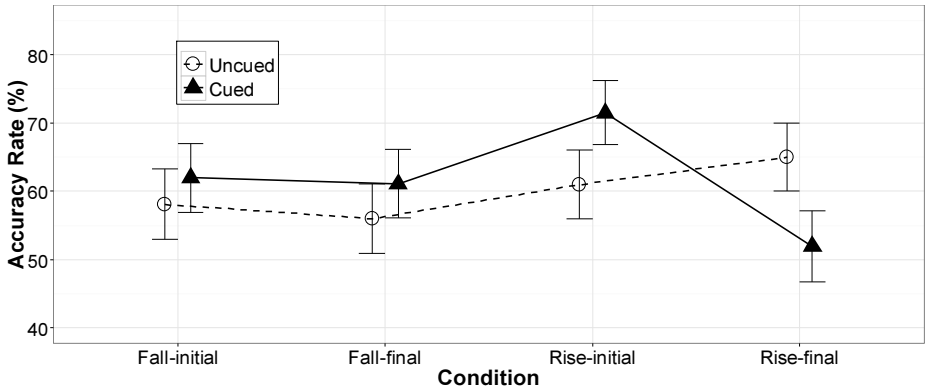


Figure 2. *Mean accuracy rates of each of the cued conditions, calculated for the words that were given the prosodic cue corresponding to their condition and for those that were not. The error bars represent 95% confidence intervals.*

[1] Kim, S. 2003. The role of post-lexical tonal contours in word segmentation. In M. J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences*, (Barcelona). Adelaide Causal Productions.

[2] Welby, P. 2007. The role of early fundamental frequency rises and elbows in French word segmentation. *Speech Communication*, 49(1), 28-48.

[3] Warner, N., Otake, T., & Arai, T. 2010. Intonational structure as a word-boundary cue in Tokyo Japanese. *Language and Speech*, 53(1), 107-131.

[4] Cutler, A., & Norris, D. 1988. The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, 14(1), 113-121.

[5] Pan, H.-H. 2007. Initial strengthening of lexical tones in Taiwanese Min. In C. Gussenhoven & T. Riad (Eds.), *Tones and Tunes: Typological Studies in Word and Sentence Prosody*. Berlin: Mouton de Gruyter, 271–292.

[6] Bodman, N. 1955. *Spoken Amoy Kokkien*. Kuaka Lumpur: Government, Federation of Malaya.

Tuesday, 12:00 - 12:30
Oral Session 4, #3

Phonetic cues in French prosodic boundaries and infant prosodic processing

Sandrien van Ommen[1], Natalie Boll-Avetisyan[2], Saioa Larraza[3], Caroline Wellmann[2], Ranka Bijeljac-Babic[1], Barbara Höhle[2], Thierry Nazzi[1]
1 CNRS, Université Paris Descartes, 2 Universität Postdam, 3 University of the Basque country

Prosody has been hypothesized to be useful early on in language acquisition due to its (imperfect) relation to syntactic structure. Indeed, several studies have demonstrated that infants with distinct language backgrounds are able to perceive the different cues marking prosodic boundaries as early as at 6 months of age (e.g. Nazzi et al., 2000). However, the perception of boundary cues seems to undergo developmental changes: some evidence suggests that infants' sensitivity to prosodic boundaries develops from a general reliance on combined cues to a language-specific cue-weighting (e.g. Johnson & Seidl, 2008; Wellmann et al., 2012).

However, differences in the types of stimuli and procedures used in previous studies make language-specific differences difficult to interpret. Therefore, we used a close French replication of Wellmann et al. (2012)'s German study, to test French-learning infants' sensitivity to a major prosodic boundary (after the second name or not of short utterances: [Loulou et Manou][et Nina],[Loulou et Manou et Nina]). An acoustic analysis of German and French pronunciations of this phrasal template (n=2880; 12 speakers per language) using Random Forest Classification (Breiman, 2001) showed that pausing and pre-boundary pitch movement have a high discriminatory potential for these boundaries, both in German and in French. The perceptual relevance of these cues for infant language development is assessed with cue elimination in consecutive behavioral experiments.

Experiment 1 tested 40 French-learning 6-month-olds' and 40 8-month-olds' ability to distinguish the phrases with or without the natural internal prosodic boundary. In Experiment 2, the internal boundary was realized by pitch and final lengthening only. Using the head-turn preference procedure, infants were familiarized to French phrases either containing a boundary or not. In a consecutive test phase, infants were exposed to both types of phrases to assess discrimination.

A linear mixed model analysis of looking times revealed a significant interaction of age, experiment and condition ($t=2.758$, $p=.006$); planned comparisons showed it was due to a familiarity preference for 8-months-olds in Experiment 1 ($z=2.121$, $p=.034$) and no significant preferences in the other groups. A control model, testing the effect of stimulus instead of familiarity, revealed an overall trend for a preference for the with-boundary condition ($t= 1.941$, $p=.052$) and a significant four-way interaction of age, experiment, stimulus and familiarization ($t=2.758$, $p= .006$); planned comparisons showed it was due to a significant preference for with-boundary stimuli in French-learning 6-month-olds familiarized with the no-boundary stimuli ($z=2.219$, $p=.027$) and no significant preferences in the other groups.

We conclude that French-learning 8-month-olds can discriminate prosodic boundaries when all cues spontaneously produced by a French speaker are present, interpretable as linguistic discrimination, while failing to discriminate the same boundaries when they were realized by pitch and final lengthening only. This behavior is in contrast with German-learning 8-month-olds who were able to discriminate prosodic boundaries realized by pitch and final lengthening (Wellmann et al, 2012). French-learning 6-month-olds' preference for novel natural with-boundary stimuli is more straightforwardly interpretable as an acoustic interest which is in line with German-learning 6-month-olds' behavior (Wellmann et al, in preparation). Follow-up cross-linguistic experiments testing French-learning infants with German stimuli and German-learning infants with French stimuli should clarify whether the cross-linguistic differences found so far can be attributed to stimuli-specific boundary salience, or to language-specific cue-weighting in processing.

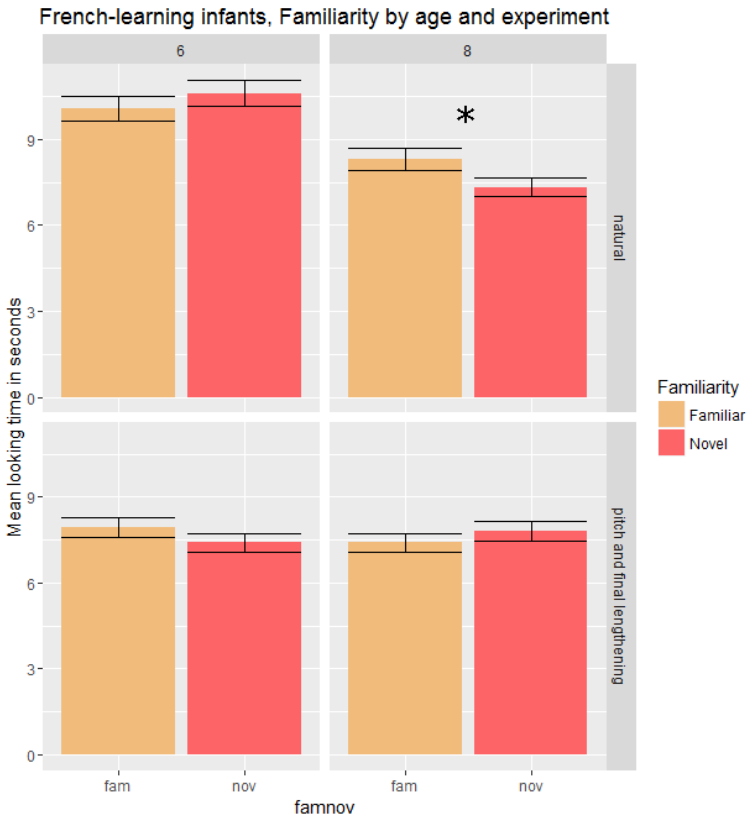


Figure 1. *Familiarity by age and experiment in French-learning infants*

[1] Nazzi, T., Jusczyk, P. W., and Johnson, E. K. (2000). Language discrimination by English-learning 5-month-olds: Effects of rhythm and familiarity. Journal of Memory and Language, 43(1):1–19.
[2] Johnson, E. K. and Seidl, A. (2008). Clause segmentation by 6-month-old infants: A cross-linguistic perspective. Infancy, 13(5):440–455.
[3] Wellmann, C., Holzgrefe, J., Truckenbrodt, H., Wartenburger, I., and Höhle, B. (2012). How each prosodic boundary cue matters: Evidence from German infants. Frontiers in Psychology.
[4] Hunter, M. A. and Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. Advances in infancy research.
[5] Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.

Tuesday, 13:15 - 14:15
Poster Session 3, #1

Do Spaniards speak faster than Mexicans? Studying Spanish rhythm in natural speech

Fabian Santiago[1] & Paolo Mairano[2]
[1]*Université de Paris 8 & Sorbonne Nouvelle,* [2]*University of Warwick*

Goals. Many studies have analysed differences in temporal patterns across different varieties of English [1], French [2], Italian [3] and many other languages, but only few studies have compared Spanish varieties. These have pointed out that articulation rate is affected by extra-linguistic factors such as age and professional profile of speakers, but not by different dialects (e.g. [4] compared Costa Rican and Barcelonan Spanish). This study compares the temporal patterns of Castilian Spanish (Madrid) and Mexican Spanish (Mexico City). Our first goal is to investigate variation in articulation rate (AR) across these varieties, also considering the effect of other factors such as gender and speech style. Our second goal is to examine whether the differences in AR affect the compression of the vocalic space similarly. Unstressed vowels tend to be centralized, whereas stressed ones do not; we hypothesize that the group associated to a faster articulation rate should compress the vocalic space to a greater extent.

Methods. 22 monolingual speakers (12 Mexicans and 10 Spaniards, age range = 21-38 years, M = 29.1, SD = 4.3) were recorded for a total of 5.75 hrs of speech. Participants performed a semi-directed interview, the description of a painting, and a reading task. The data was orthographically transcribed and segmented into phones, syllables and words with EasyAlign [5]; the segmentation was then entirely checked manually. The corpus amounts to 45k words and 7,977 IPUs (Inter-Pausal Units, see Fig. 1). For the present analysis, we excluded all IPUs in which one or more word(s) contained disfluencies, hesitations, filled pauses, etc. The remaining 4,618 IPUs consisted of speech chunks characterised by continuous phonation without noticeably lengthened words. We used two metrics. The first one is the AR measured in number of syllables per second (excluding pauses). The second one is a measure of the vowel space. We automatically extracted F1 and F2 values at the mid-point of all the vocalic segments. Formant values were converted into Barks for normalization. We measured each speaker's vowel space in terms of 'degree of dispersion' following [6]: (i) we identified the gravity centre of the acoustic space by averaging F1 and F2 values of all vowels produced by each speaker; (ii) we computed the Euclidean distances from the gravity centre to each vowel's position within the vowel chart, then averaged them for the 5 peripheral Spanish vowels, separately for the two conditions stressed/unstressed; (iii) we summed the resulting averaged Euclidean distances for the 5 peripheral Spanish vowels in order to evaluate to what extent the size of the vowel space was affected by the presence/absence of lexical stress. The data was analysed building linear mixed models with random intercepts and slopes for participants.

Results & Discussion. Our study shows that Castilian speakers articulate faster than Mexicans (Fig. 2a). Speech style affects both groups in a similar way: AR increases in reading but decreases in spontaneous speech (Fig. 2a). Gender affects the AR, but not in the same way across the two varieties analysed: on average, male participants speak faster than female participants (Fig. 2b). However, Castilian female speakers read faster than Mexican female speakers, whereas male speakers of the two dialectal varieties show similar ARs (Fig. 2c). In spontaneous speech, female speakers articulate faster than male speakers within the Castilian group, whereas the opposite pattern is observed for Mexican speakers (Fig. 2d). As for vocalic compression, we observed that Spanish speakers do not tend to compress their vowel space as much as Mexican speakers do (Fig. 3), even though they articulate faster. This is contrary to our initial hypothesis. This observation can perhaps be explained by differences in the rhythm typology of these two varieties: Mexican Spanish could be more stress-timed than Castilian Spanish so that, despite a slower AR, vowel reduction applies more widely in the former.

References.

[1] Jacewicz, E., Fox, R. A., & Wei, L. (2010). Between-speaker and within-speaker variation in speech tempo of American English. *Journal of the Acoustical Society of America*, 128, 839-350.

[2] Schwab, S. & Avanzi, M. (2015). Regional variation and articulation rate in French. *Journal of Phonetics,* 48, 96-105.

[3] Russo, M. & Barry W.J. (2004). Interaction between Segmental Structure & Rhythm. A look at Italian Dialects and Regional Standard Italian. *Folia Linguistica*, 38, 277-296.

[4] Schwab, S. (2015). Las variables temporales en el español de Costa Rica y de España: un estudio comparativo. *Filología y Lingüística,* 41 (1), 127-139.

[5] Goldman, J.Ph. (2011). EasyAlign: an automatic phonetic alignment tool under Praat. *Proceedings of InterSpeech,* Firenze, Italy.

[6] Gendrot, C. & Adda-Decker, M. (2007). Impact of duration and vowel inventory size on formant values of oral vowels: an automated formant analysis from eight languages. *International Conference of Phonetic Sciences 2007*, 1417-1420.

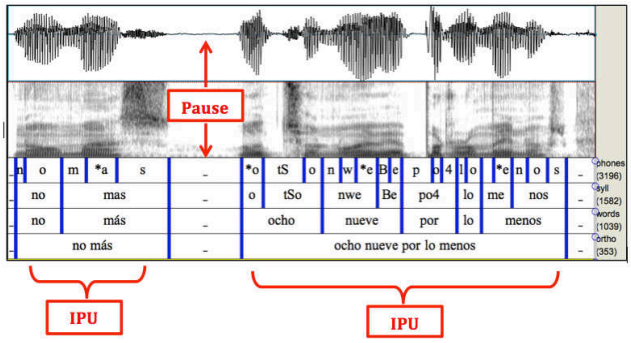


Figure 1

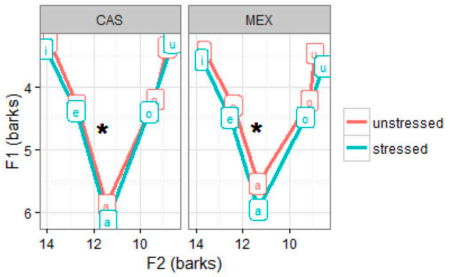


Figure 3

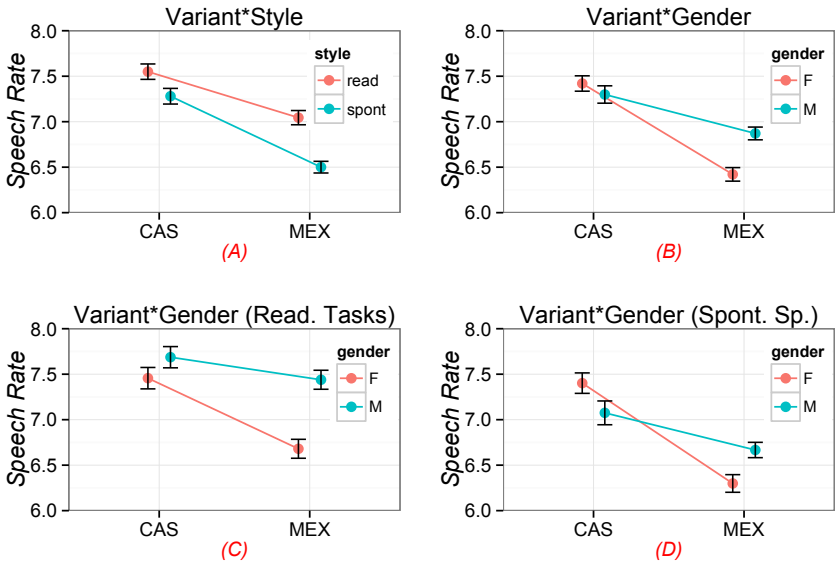


Figure 2

The prosody of verb-first constructions in German: A comparison of information-seeking questions, rhetorical questions and exclamatives

Daniela Wochner & Nicole Dehé
University of Konstanz

Sentence form does not fully predict pragmatic function. For instance, German verb-first sentences may either express neutral information-seeking questions (lack of information), rhetorical questions (belief) or exclamatives (surprise). We investigate prosodic differences as a function of pragmatic function in verb-first constructions.

A study by [1] already revealed some prosodic characteristics of verb-first (V1) rhetorical questions (RQs) in comparison to information-seeking questions (ISQs): lower initial pitch, longer utterance duration, softer voice in the verb in V1 position and a difference in boundary tones - RQs were realized more often with a final plateau contour L* H-%, ISQs with a low-rising contour L* H-^H%. Interestingly, some of the properties of rhetorical questions have been associated with the expression of surprise ([2], [3] & [4]), in particular breathier voice, lower initial pitch and segmental lengthening. The similarities of the prosodic profiles of the different utterance types (RQs and surprised utterances) beg the question how specific the prosodic cues are for certain pragmatic functions.

We recorded eight German participants (4 female) who produced RQs and string-identical exclamatives (EX) and ISQs, which were embedded in contexts (see Table 1). Pragmatic function was manipulated within-subjects. The target utterances were annotated syllable-by-syllable, pitch accents and boundary tones were annotated following GToBI ([5]) and pitch values were extracted at f0 turning points. Vocal effort was operationalized as H1*-A3*. In order to analyze which of the acoustic measures explain best the differences between the three pragmatic types, they were entered in a multinomial linear regression model with CONDITION (ISQ, RQ and EX) as predictor variable. For the significant predictor variables (see Figure 1), we calculated linear mixed effect regression models with CONDITION (ISQ, RQ and EX) as fixed factors and SUBJECT and ITEM as crossed random factors. Pitch onset, the duration of other stressed syllables, H1*-A3* and f0-range in the light verb were not significant.

The results show that RQs and EXs share prosodic features, which are, however, realized to varying degrees. Both RQs and EXs may be produced with a falling or rising intonation contour, however, RQs were produced more often with a final rise (RQ: 84%, EX: 25%), which is also larger in expansion than in EXs (RQ: 10.6 semitones, EX: 4.8st). ISQs are mainly rising and have the largest pitch offset (12.9st). Also, RQs and EXs show more prenuclear pitch accents (RQ: 84%, EX: 98%, ISQ: 48%), which also had a higher pitch range than ISQs (EX: 3.7st, RQ: 2.2st, ISQ: 1.8st; Figure 2). Duration measurements of the verb in V1 reveal that it is shortest in ISQs, followed by RQs and EXs (ISQ: 134ms, RQ: 145ms, EX: 152ms), but only the comparisons involving ISQs become significant.

We assume two scales for the classification of the pragmatic function of verb-first sentences with regard to prosody: the first scale is an interrogativity scale ranging from high interrogativity (ISQs) to low/ no interrogativity (EXs). This scale is phonetically and phonologically manifested in the pitch excursion of the offset of the utterances and in the choice of the type of boundary tone. The second scale constitutes a range from highly expressive and emphatic (EXs) to neutral and non-emphatic (ISQ) and is visible in the durational values of the verb and the pitch range in the prenuclear field. In both scales rhetorical questions are placed between the two extremes (ISQs, EXs), which we interpret as being due to the mismatch between form (questions) and functions (assertive-like), which has been reported for RQs (c.f., [6] & [7]).

Literature

[1] Wochner, Daniela, Jana Schlegel, Nicole Dehé & Bettina Braun (2015). The prosodic marking of rhetorical questions in German. *Proceedings of the 16th Annual Conference of the International Speech Communication Association (INTERSPEECH 2015), 6-10 September, Dresden, Germany*.

[2] Kehrein, Roland (2002). The Prosody of Authentic Emotions. *Proceedings of Speech Prosody, April 11-13 2002, Aix-en-Provence, France,* pp. 423-426.

[3] Kohler, Klaus & Oliver Niebuhr (2007). The phonetics of emphasis. *Proceedings of the 16th ICPhS, 6-10 August, Saarbrücken, Germany,* pp. 2145-2148.

[4] Niebuhr, Oliver, Julia Bergherr, Susanne Huth, Cassandra Lill & Jessica Neuschulz (2010). Intonationsfragen hinterfragt – die Vielschichtigkeit der prosodischen Unterschiede zwischen Aussage- und Fragesätzen mit deklarativer Syntax. *Zeitschrift für Dialektologie und Linguistik* 11, pp. 304-346.

[5] Grice, Martine & Stefan Baumann (2002). Deutsche Intonation und GToBI. *Linguistische Berichte* 191, pp. 267-298.

[6] Han, Chung-hye (2002). Interpreting interrogatives as rhetorical questions. *Lingua* 112(3), pp. 201-229.

[7] Rohde, Hannah (2006). Rhetorical questions as redundant interrogatives. *San Diego Linguistic Papers* (2), pp. 134-168.

Appendix

a. Information-seeking question (expresses: lack of information)	Du möchtest Lene, der Tochter einer Freundin, Buntstifte zum Geburtstag schenken. Du weißt jedoch nicht, ob die Kleine schon malen kann oder nicht. Du sagst zu deiner Freundin: **Kann die Lene malen**
b. Exclamative (expresses surprise)	Als du mit einem Freund zum ersten Mal das Atelier eurer Freundin Lene besuchst, bist du überrascht, wie gut sie malen kann. Du sagst zu deinem Freund: **Kann die Lene malen**
c. Rhetorical question (expresses belief)	Deine Freundin möchte, dass eure Mitbewohnerin Lene ein Bild für einen Freund malt. Doch es ist überall bekannt, dass Lene überhaupt nicht malen kann. Du sagst zu deiner Freundin: **Kann die Lene malen**

Table 1. Context examples with the target sentence *Kann die Lene malen*.

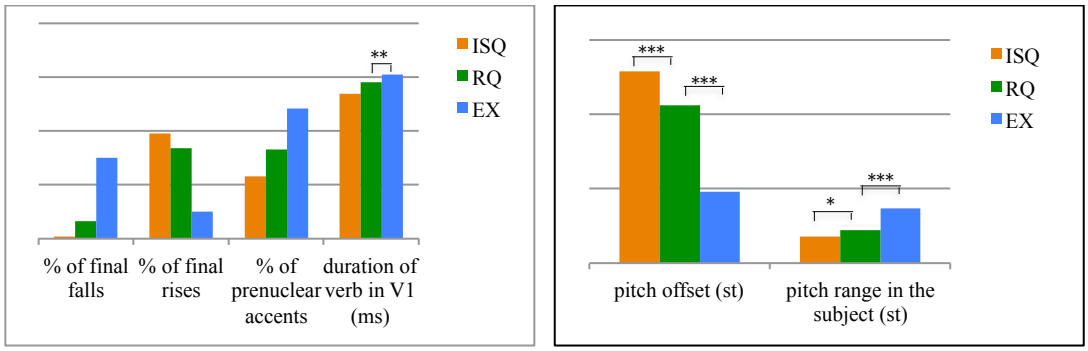


Figure 1. Distribution of variables in the three sentence types (*: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$).

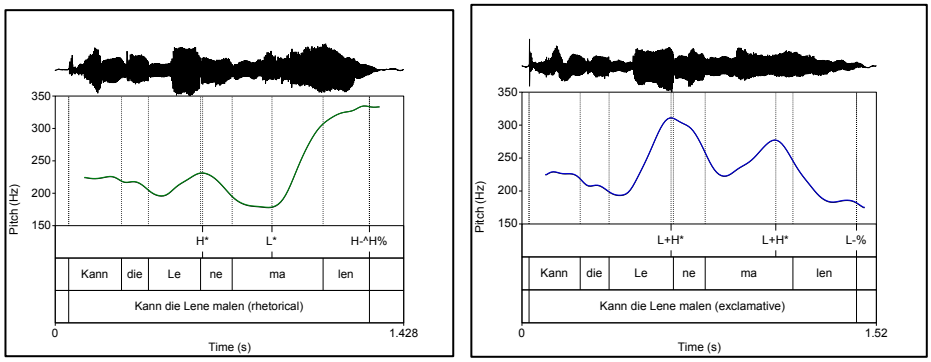


Figure 2. Example of *Kann die Lene malen* in rhetorical and exclamative condition (produced by female speaker).

Perception of the Downstepped Final Accent in French

Rogelio Méndez and Corine Astésano
Octogone-Lordat (EA 4156), Université de Toulouse 2 (UT2), France

French accentuation is not lexically distinctive, contrary to most languages. It is traditionally viewed as a 'language without accent' [1] or primarily a boundary language [2], because phrase-final accent (hereafter 'FA') is co-occurrent to the boundary tone, which makes accent and boundary phonetically difficult to disentangle. In this view, intonation would supersede prominence (stressed-final syllable) at hierarchical levels above the *ap* [1; 3]. However, these descriptions are essentially based on *tonal* accounts of regrettably syncretic tonal events. Some approaches [4; 6] argue that other metrical cues should be taken into account alongside tonal parameters (duration and metrical weight). They argue that perception should be used as an interface between phonetic information and phonological representations, in order to account for only those *metrical* events that are actually used by listeners.

A number of studies have shown that French not only has a primary FA (canonically characterized by both a *f0* rising tone and a longer duration), but also a secondary Initial Accent (hereafter 'IA'; [3; 4]), perceived as stronger than FA [6, 9]. However, no experimental account has yet been proposed for a less canonical type of FA realization: the utterance-medial FA characterized by a *f0* falling tone. It was first described by Delattre [7] in some instances of syntactic dependencies at the level of the minor phrase (*ap*). Some authors have recently proposed to qualify it as a 'Downstepped Final Accent' (!H*: [8]; hereafter 'DFA'). Both accents (FA and DFA) are aligned with stressed syllables. We conducted two perceptual experiments on 20 French native listeners, using a specific syntactic structure where two Nouns are qualified by an Adjective, in all-focus sentences (as in [(les bonimenteurs)*ap* (et les baratineurs)]*ip*[(fabuleux)]; [5]) (Figure 1), in order to determine: (a) whether DFA is perceived as prominent (*i.e.* stressed) in the same way as FA, and more so than unstressed syllables (hereafter 'UN'); and (b) if boundaries are perceived as distinct phenomenon not only from FA (as in [6; 9]) but also from this less canonical DFA. Subjects were presented with 32 French sentences taken from 4 speakers, chosen to equally exhibit FA and DFA at our two levels of prosodic hierarchy (*ap* and *ip*). Participants were asked to indicate, on a scale from 0 (none) to 3 (strong), the degree of prominence perceived on each syllable (prominence task) and the degree of boundary between each word (boundary task). Listeners were equally split into two groups, in order to counterbalance the tasks (prominence and boundary) presentation order.

Results indicate the following prominence strengths: IA = FA > DFA > UN (Figure 2.a). Firstly, IA and FA are perceived as equally strong prominences, while FA is stronger than DFA ($p<.001$). Syllables carrying a DFA are perceived as more prominent than unstressed syllables ($p<.0001$). This result is a first experimental step towards the inscription of DFA in the French metrical structure, and it shows that even when these syllables are not marked tonally by a canonical *f0* rise, listeners tend to perceive them as metrically strong, indicating that they are encoded at a phonological level. Secondly, results for the boundary task show that final prominences are perceived differently according to the level of boundaries at which they occur (Figure 2.b). Main effects show that both FA and DFA are perceived as weaker at a*p* boundaries than *ip* boundaries; post-hoc tests show that FA at *ip* boundaries are perceived weaker than DFA at the same level ($p<.001$), whereas both are perceived as equally strong at *ap* boundaries ($p=0.0057$). Thus, our results confirm previous claims [6; 9] according to which French listeners are not stress-deaf and are able to perceive final prominences as (partially) distinct phenomena from boundaries, because their *metrical* weight survives higher prosodic levels. Our results however indicate that phonetic variations modulate the phonological weight of FAs and the relationship between metrical weight and boundary marking: DFA reinforces the perception of stronger *ip* boundaries.

(1)

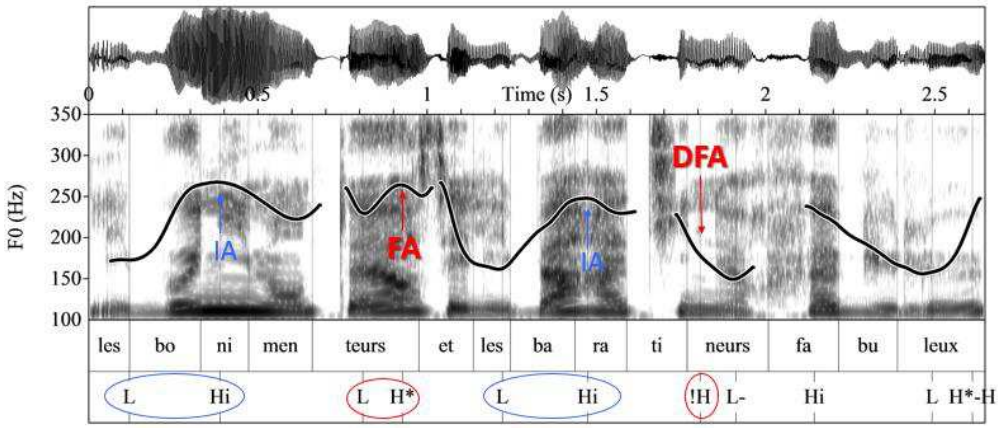


Figure 1. *French prominences IA (Hi), FA (H*) and DFA (!H) in the utterance 'Les bonimenteurs et les baratineurs fabuleux'* [*'The fabulous smooth-talkers and rug-peddlers'*].

(2)

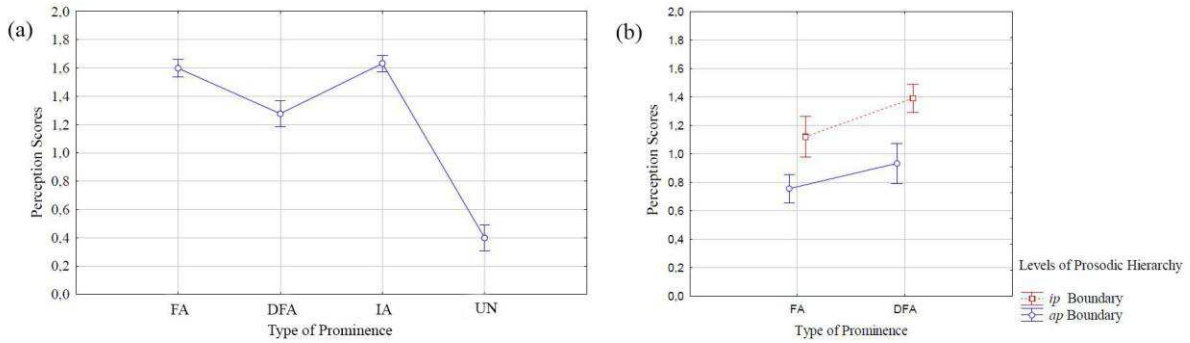


Figure 2. (a) *Perception of French prominences (IA = FA > DFA > UN), on a scale from 0 to 3.* (b) *Perception of FA (H*) and DFA (!H) associated to ap and ip boundaries.*

[1] Rossi, M. (1980). Le français, langue sans accent ? In I. Fónagy & P. Léon (Eds.). *L'accent en français contemporain* (pp. 13–51). Ottawa: Didier.

[2] Vaissière, J. (1990). Rhythm, accentuation and final lengthening in French. In J. Sundberg, L. Nord & R. Carlson (Eds.). *Music, language, speech and brain.* (pp. 108- 120). MacMillan Press.

[3] Jun, S.-A., & Fougeron, C. (2000). A Phonological model of French intonation. In A. Botinis (Ed.). *Intonation: Analysis, Modeling and Technology* (pp. 209–242). Dordrecht: Kluwer Academic Publishers.

[4] Di Cristo, A. (2000). Vers une modélisation de l'accentuation du français (seconde partie). *Journal of French Language Studies*, 10(1), 27–44.

[5] Astésano, C., Bard, E. G., & Turk, A. (2007). Structural influences on initial accent placement in French. *Language and Speech*, *50*(3), 423–446.

[6] Astésano, C., & Bertrand, R. (2016). Accentuation et niveaux de constituance en français: enjeux phonologiques et psycholinguistiques. *Langue française*, (2016/3 ; n°191 : *La Prosodie du français : accentuation et phrasé*), 11-30.

[7] Delattre, P. (1966). Les dix intonations de bases du français. *French Review*, *40*(3), 1–14.

[8] Welby, P., Bertrand, R., Portes, C., & Astésano, C. (2016). Realization of the French initial accent: Stability and individual differences. *TIE Conference 2016 – 7th conference on Tone and Intonation in Europe (TIE)*. Canterbury. 1-3 September 2016.

[9] Astésano, C., Bertrand, R., Espesser, R., & Nguyen, N. (2012). Perception des frontières et des proéminences en français. *JEP-TALN-RECITAL 2012* (pp. 353–360). Grenoble.

The Role of Perception in Learning Vowel Nasalization

Kim Strütjen, Dinah Baer-Henney, Peter Indefrey and Ruben van de Vijver
Heinrich-Heine University Düsseldorf

Recent research has provided arguments in favor of the hypothesis that phonetically based phonological rules are learnt more easily than rules that are not [1], [2] or less so [3], [4]. Phonetically based rules facilitate either production or perception. The present study investigates whether such a bias influences the acquisition of vowel nasalization in relation to vowel height. Low oral and nasalized vowels are acoustically more similar than non-low oral and nasalized vowels [5], [6]. Moreover, high nasalized vowels are easily distinguished from high oral vowels in perception experiments [7], [8].

To investigate whether perception may be used to predict how German adults learn a vowel nasalization rule, we ran a perception experiment and an artificial language learning experiment. The results of the perception experiment were modeled in a Maximum Entropy (MaxEnt) grammar [9], [10] which was used to predict the outcomes of the learning experiment.

Our participants had no experience with nasalized vowels [11]. In the perception experiment 28 native German participants listened to vowels in the presence of a masking noise and were forced to identify the vowels as one of six oral and nasalized vowels differing in height. We found that low nasalized vowels were more likely to be confused with low oral vowels and that mid nasalized vowels and high nasalized vowels are better distinguishable from mid and high oral vowels. We used the results of this experiment to create a MaxEnt grammar. Our aim was to compare the grammars' predictions with the results of the artificial language learning experiment.

In the learning experiment 60 participants were exposed to an alternation in which vowels are nasalized before nasals but not before laterals. Participants were trained in three different experimental groups with items containing either high [i], mid [ɛ] or low [a] vowels. In the subsequent forced-choice-task they were tested on items containing all three vowel heights and were forced to make grammaticality judgments about oral and nasalized vowels either in agreement or not in agreement with the language they had been exposed to. The results show a learning advantage for non-low nasalized vowels over low nasalized vowels: Learners from the high and mid group acquire the pattern in the trained height more readily than learners from the low group (Figure 1). The advantage for high and mid vowels is also present in the generalization behavior: Nasalization is generalized more often to untrained higher vowels than to untrained lower vowels. Learners from the low group do not distinguish between high and mid vowels (Figure 1).

We provided three different MaxEnt grammars—an [a]-grammar, an [ɛ]-grammar and an [i]-grammar—which we derived from our perception experiment with the same inputs our participants had been exposed to in training. We found that the grammars' predictions match our experimental results well, as is illustrated in Figure 1. The best performance for both our grammars and our participants in the learning experiment is observed in the trained conditions. With the exception of the [a]-grammar and the participants trained with [a] vowels, we observe a generalization to untrained higher vowels both for our grammars and for our participants.

Thus, a grammar based on perception can explain the asymmetrical results in learning vowel nasalization and indicates an active role of phonetics in phonology. This shows that learners are phonetically biased in learning phonological patterns.

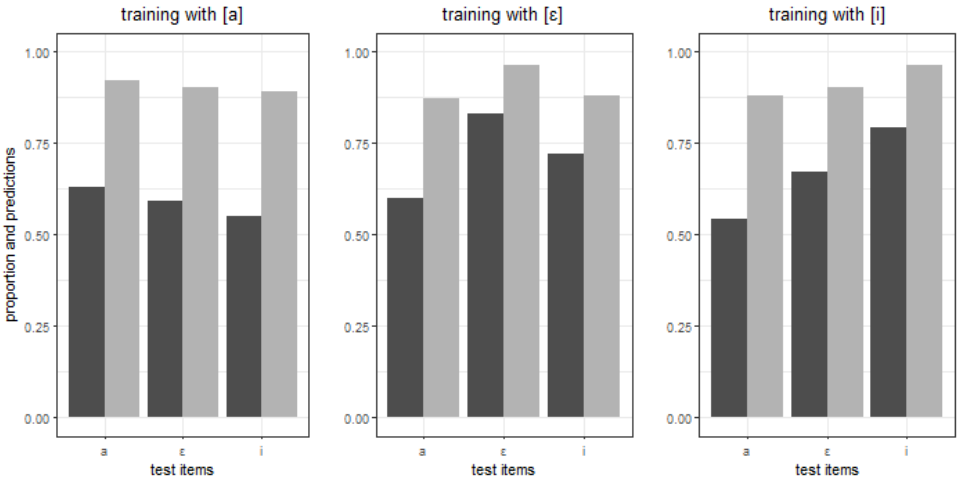


Figure 1. *Comparison of experimental results (black bars) and MaxEnt models' predictions (grey bars).*

[1] Baer-Henney, D., Kügler, F. & van de Vijver, R. 2014. The Interaction of Language-Specific and Universal Factors during the Acquisition of Morphophonemic Alternations with Exceptions. *Cognitive Science*, 39(7), 1537-1569.

[2] van de Vijver, R. & Baer-Henney, D. 2014. Developing Biases. *Frontiers in Psychology*, 5(634), 1-8.

[3] Finley, S. 2012. Typological Asymmetries in Round Vowel Harmony: Support from Artificial Grammar Learning. *Language and Cognitive Processes*, 27(10), 1550-1562.

[4] Wilson, C. 2006. Learning Phonology with Substantive Bias: An Experimental and Computational Study of Velar Palatalization. *Cognitive Science*, 30(5), 945-982.

[5] Ohala, J. J. 1975. Phonetic Explanations for Nasal Sound Patterns, In Ferguson, C. A., Hyman, L. M. & Ohala, J. J. (Eds.), *Nasálfest: Papers from a Symposium on Nasals and Nasalization*, Stanford: Language Universals Project, 289-316.

[6] Schwartz, M. F. 1968. The Acoustics of Normal and Nasal Vowel Production. *The Cleft Palate Journal*, 5(2), 125-140.

[7] Bond, Z. S. Identification of Vowels Excerpted from Neutral and Nasal Contexts. *Journal of the Acoustical Society of America*, 59(5), 1229-1232.

[8] House, A. S. & Stevens, K. N. 1956. Analog Studies of the Nasalization of Vowels. *Journal of Speech and Hearing Disorders*, 21(2), 218-232.

[9] Goldwater, S. & Johnson, M. 2003. Learning OT Constraint Rankings Using a Maximum Entropy Model. In Spenador, J., Eriksson, A. & Dahl, O. (Eds.), *Proceedings of the Stockholm Workshop on Variation within Optimality Theory* (Stockholm), pp. 111-120.

[10] Hayes, B. & Wilson, C. 2007. A Maximum Entropy Model of Phonotactics and Phonotactic Learning. *Linguistic Inquiry,* 39(3), 379-440.

[11] Wiese, R. 1996. The Phonology of German. Oxford: Oxford University Press.

On the realization of bouletic bias in German questions

Sophie Egger, María Biezma, Bettina Braun
University of Konstanz

In everyday life we use subtle ways to communicate desires, often without explicitly saying so. One way to indirectly utter these desires is by asking questions. Questions with this additional non-truth-conditional aspect are referred to as *biased* [1]: they are not plainly information seeking but express an attitude towards one of the possible answers, e.g. a wish or desire in a *bouletic bias* [1-5]. We hypothesize that speakers successfully convey their desires when subliminally expressing them in a biased question, given that interlocutors typically comprehend them. Therefore, we investigated how speakers use morphosyntactic and prosodic cues to encode bouletic bias in German questions, specifically looking at polar (PolQs) vs. alternative questions (AltQs). Since PolQs highlight one particular alternative from a set of propositions [6, 7], we expect them to be more appropriate to mark bias. In AltQs, on the other hand, both alternatives are equally epistemically available, which makes them suitable to offer an unbiased (i.e., neutral) choice [6]. We hence predict that speakers chose more positive PolQs in biased and more AltQs in unbiased contexts.

In a production experiment, we used 32 situational contexts, evoking either a biased (16 contexts) or a neutral question (16 contexts, within-items design; see Table 1). They were presented together with either a PolQ or an AltQ (manipulated between-items), resulting in 32 trials per participant. Each trial started with a context displayed on screen. Upon pressing a button, participants saw a target question (PolQ or AltQ; see Table 1), which they were asked to produce (part 1). After another button press, they were given the possibility to rephrase the question in a way that seemed most natural (part 2). In the analysis we will focus on part 2, since these productions will directly reveal the morphosyntactic structure preferred for biased and neutral questions, respectively, as well as prosodic differences between the two.

We recorded 16 German speakers (Ø = 23.3 years, 2 male). They produced 375 target questions (201 biased, 174 neutral) in part 2. The productions were coded for syntactic type (AltQ, PolQ, tag-question, *wh*-question, other). Participants predominantly produced PolQs in the biased condition (74%) and AltQs in the neutral condition (69%); see Figure 1. The question types presented in part 1 were changed in 70% of the biased contexts from AltQ to PolQ, and in 67% of the neutral contexts from PolQ to AltQ, showing strong preferences for particular question types according to context.

For the prosodic analysis, we manually annotated a subset of 114 productions from part 2 according to GToBI [8]. The subset consisted of all non-prototypical choices (30 biased AltQs, 27 neutral PolQs), as well as 30 neutral AltQs and 27 biased PolQs (randomly selected). Results showed that AltQs were generally produced with a final low plateau (L*L-%) in both conditions (biased: 75%, neutral: 83%,), while in PolQs, the most frequent nuclear tone was a final rise (L* H-^H%; biased: 50%, neutral: 45%). In neutral PolQs, there were more final rises ending in a high plateau, than in biased PolQs (L*/H* H-% biased: 18%, neutral: 32%). For a preliminary phonetic analysis, we extracted the f0 range (difference in semitones between L and H of the final rising and falling movement). Results revealed that PolQs are produced with a larger pitch range in the biased condition (10.2st) than in the neutral condition (7.7st; $p < 0.05$). Biased AltQs, on the other hand, are produced with a smaller pitch range (6.7st) than their neutral counterparts (8.7st, $p = 0.08$).

Our findings corroborate the assumption that positive PolQs tend to convey a bias [6, 7], while AltQs function as neutral questions more readily [6]. There appear to be only minor differences in the preferred intonational realization across conditions. Instead, speakers use a decreased pitch range to compensate for the non-prototypical morphosyntactic structure when producing AltQs in biased contexts and PolQs in neutral contexts. However, we leave key phonetic differences (e.g., voice quality, slope, peak alignment, durations of alternatives) for future research.

Neutral condition	Biased condition
You and your brother both want some ice cream after dinner. You see that there is one chocolate and one strawberry ice cream left in the freezer. You like both just the same, but you **don't know** which one your brother prefers. Therefore you ask him…	You and your brother both want some ice cream after dinner. You see that there is one chocolate and one strawberry ice cream left in the freezer. You like the strawberry ice cream better and you **hope** that your brother wants chocolate. You ask him…
Speaker intention (not on display)**:**	
I want to know whether you want the chocolate or the strawberry ice cream.	*I want you to take the chocolate ice cream.*
Target questions: (either PolQ or AltQ presented on screen in part 1)	
PolQ: *Do you want the chocolate ice cream? / Do you want the strawberry ice cream?* AltQ: *Do you want the chocolate or the strawberry ice cream?*	

Table 1: *Example of a neutral and biased context with speaker intention and example of a PolQ and AltQ used in both conditions (question type was manipulated between-items).*

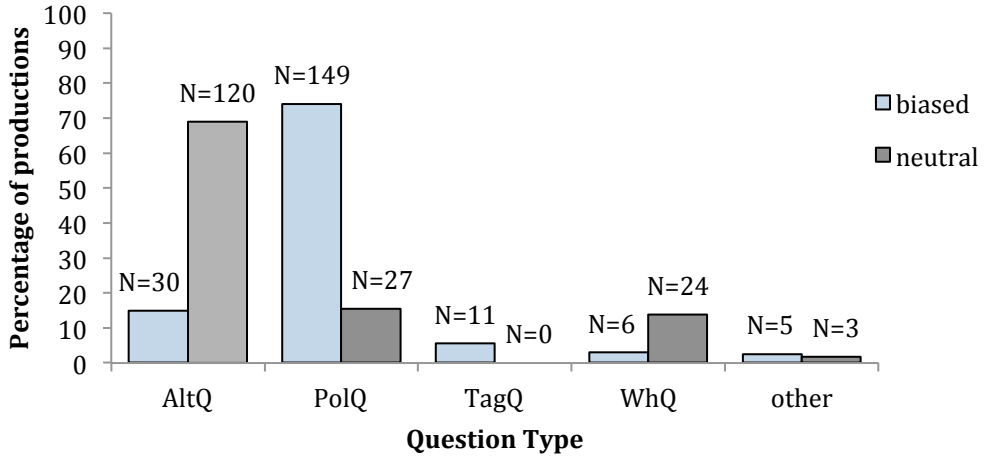


Figure 1: *Percentage of productions of each question type per condition.*

[1] Sudo, Y. 2013. Biased polar questions in English and Japanese. *Beyond expressives: Explorations in use-conditional meaning*, 275-296.

[2] Reese, B. 2007. *Bias in Questions* (PhD Dissertation). University of Texas, Austin, TX.

[3] Huddleston, R., & Pullum, G.K. 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.

[4] Reese, B. 2006. The meaning and use of negative polar interrogatives. *Empirical Issues in Syntax and Semantics*, 331-354.

[5] van Rooij, R., & Šafárová, M. 2003. On Polar Questions. Talk at the *Semantics and Linguistic Theory, 13* (Seattle, WA).

[6] Biezma, M., & Rawlins, K. 2012. Responding to alternative and polar questions. *Linguistics and Philosophy*, *35(5)*, 361-406.

[7] Huddleston, R. 1994. The contrast between interrogatives and questions. *Journal of Linguistics, 30(2),* 411-439.

[8] Baumann, S., & Grice, M. 2006. The intonation of accessibility. *Journal of Pragmatics, 38(10),* 1636-1657.

Tuesday, 13:15 - 14:15
Poster Session 3, #6

And finally, no geminates!
Pohnpeian consonantal length contrasts in initial, medial, and final position

Bradley Rentz[1] and Victoria Anderson[1]
[1]*University of Hawai'i at Mānoa*

Pohnpeian (ISO639-3 pon) is a phonetically understudied Oceanic language spoken by about 34,000 people in the Federated States of Micronesia and 12,000 in the U.S. Typologically, it is uncommon for a language to have word initial and final geminates, and even rarer to have them initially, medially, and finally, but Pohnpeian is claimed to be such a language. Only the voiced consonants /m/, /mʷ/, /n/, /ŋ/, /l/, and /r/ may occur as singletons or geminates in Pohnpeian. All of these are allowed to be geminated medially [2, 3, 4]. In addition, /m/, /mʷ/, and /ŋ/ reportedly may occur as geminates word initially (although [4] state that word initial geminates are degeminated unless a prefix is added), while /l/ and /mʷ/ reportedly may occur as geminates word finally. This study investigates duration differences among geminate/singleton pairs in these positions to probe the nature of this contrast.

Three male and two female native speakers were recorded in Honolulu, Hawai'i saying phonetically controlled wordlists in three sentence frames: (1) ___ irail kin inda *('___ they always say';* (2) irail kin inda ___ *('They always say ___')*; and (3) irail kin inda ___ nimenseng *('They always say ___ in the morning')*. The target words included (near) minimal pairs like /mʷmʷus/ *'to vomit'* versus /mʷus/ *'to move as a group'*, /kɐmʷmʷus/ *'to cause to vomit'* versus /kɐmʷus/ *'to jerk up'*, and /kull/ *'cockroach'* versus /kul/ *'to grab',* for every reported geminate/singleton contrast in the three positions. Durations of segments were measured based on time-aligned waveforms and spectrograms in *Praat* [1], as shown in Figure 1.

The results of Bayesian hierarchical linear modeling demonstrate significant differences in duration for initial and medial geminate/singleton pairs. Word final pairs, however, do not exhibit clear significant duration differences overall. Word final /mʷ/ geminate/singleton pairs approach a significant difference, but given the amount of overlap of their posterior distributions there remains a strong probability that there is no duration difference between "geminates" and "singletons" in word final position. Figure 2 depicts duration differences between geminate/singleton pairs by position in word.

For conditions where geminates and singletons did show significant differences, we found the difference to be smaller word medially than word initially. However, segments (whether geminates or singletons) had shorter overall durations in word initial position than medially or finally, while segments (whether geminates or singletons) in the latter two positions were similar to each other in duration. Word initial geminates are predicted to be on average 80.02 ms longer than singletons (i.e., a ratio of 2.02:1). Word medial geminates are predicted to be on average 45.12 ms longer than word medial geminates (i.e., a ratio of 1.45:1). The so-called word final geminates are predicted to be only 6.61 ms longer on average than singletons (i.e., a ratio of 1.05:1); not different enough to constitute a meaningful contrast. The pair /mʷ/ and /mʷmʷ/ showed substantial overlap in values in all positions, bringing the phonemic status of this particular contrast into doubt.

Overall we observed, contrary to claims of previous studies, that based on empirical measurements of duration, Pohnpeian (i) does not have clear surface word final geminates but (ii) does has surface word initial and word medial geminates.

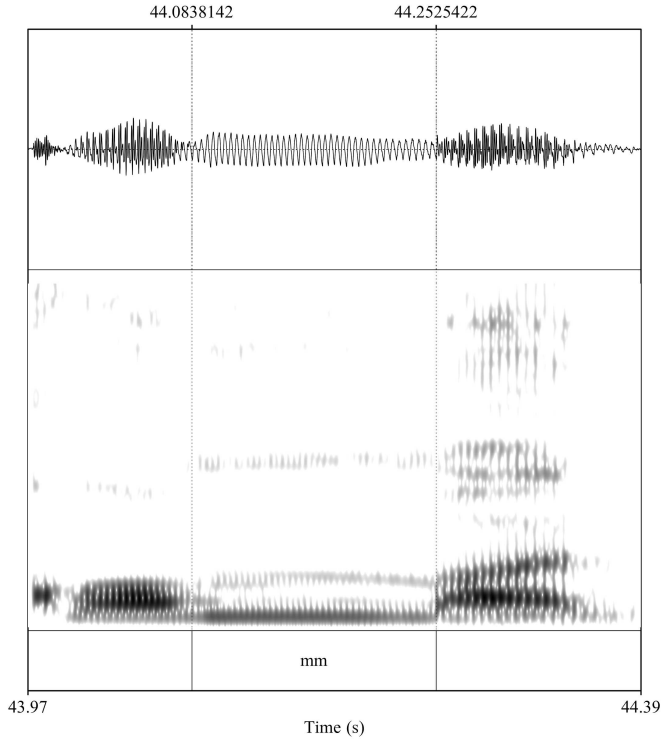


Figure 1. Measurement of /mm/ in the word /kommɔl/ 'to rest'

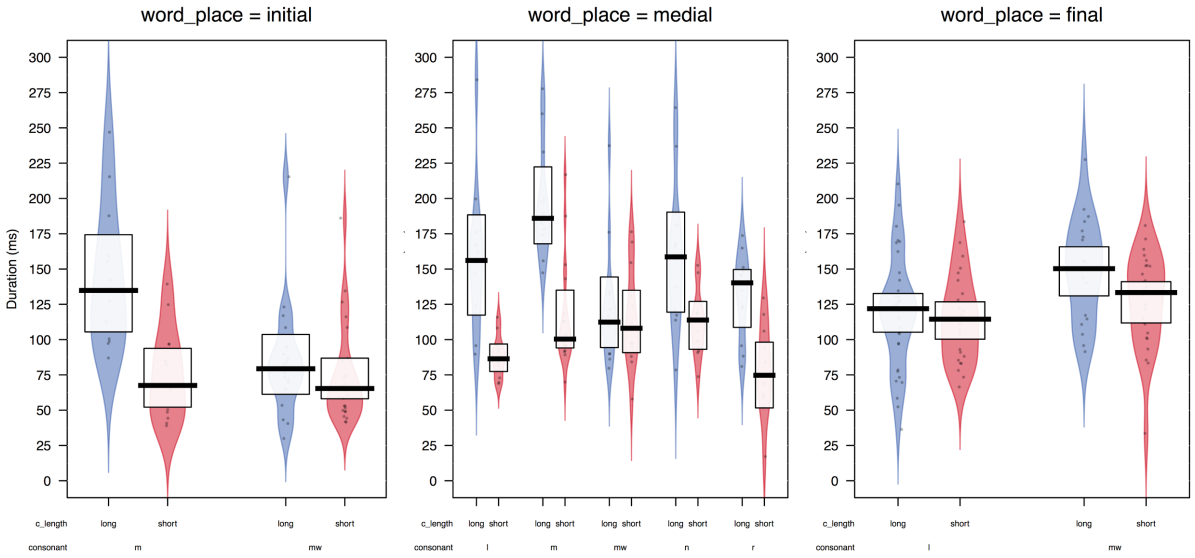


Figure 2. RDI plot of Pohnpeian geminate/singleton pairs by position in the word (Box indicates Bayesian 95% highest density interval with median bar)

[1] Boersma, Paul & David Weenink. 2015. Praat: doing phonetics by computer (version 5.4.18). http://www.praat.org.

[2] Goodman, Beverley D. 1995. *Features in Ponapean phonology*: Cornell University dissertation.

[3] McClintock, Susan. 1999. *Consonant sequence constraints in Fijian and Micronesian languages*: University of Hawai‘i at Mānoa dissertation.

[4] Rehg, Kenneth L. & Damian G. Sohl. 1981. *Ponapean reference grammar*. PALI language texts. Honolulu: University of Hawai‘i Press.

Tuesday, 13:15 - 14:15
Poster Session 3, #7

The relationship between syllable structure complexity and vowel reduction processes

Shelece Easterday
University of New Mexico

I present typological evidence that languages with different degrees of syllable structure complexity also exhibit different rates and kinds of vowel reduction processes.

[1] presents a measure of syllable structure complexity which categorizes languages according the shape and size of their canonical syllable patterns. Subsequent studies have established positive correlations between these categories, consonant phoneme inventory size, and the occurrence of elaborated consonant classes [2]. These correlations suggest that this typology is a promising avenue for uncovering crosslinguistic patterns in sound systems. A complication with this typology is that one of the categories — Complex Syllable Structure — is extremely heterogeneous, incorporating everything from biconsonantal codas to much larger consonant clusters, e.g., O'odham *ʔatʃ͡spk* [3]. In general, research on highly complex patterns like that illustrated by the O'odham example is limited and tends to focus on the problems that these rare structures pose to theoretical models of the syllable (cf. [4], [5]).

Few studies examine the motivations behind the evolution and maintenance of Highly Complex Syllable Structure (HCSS), but comparative, historical, and instrumental evidence suggests that specific kinds of vowel reduction processes contribute to the development of these structures and their characteristic phonetic features (cf. [6] for Lezgian). The current study seeks to contribute to our understanding of HCSS by investigating the typological relationship between syllable structure complexity and the prevalence and characteristics of active processes of vowel reduction in languages.

I investigate these issues in a sample consisting of 100 diverse languages distributed among four categories according to syllable structure complexity: the three categories presented in [1] and an additional category of HCSS. All reported productive phonetically-conditioned processes resulting in the deletion or reduction in duration, quality, or voicing of a vowel were collected from grammatical descriptions of the languages. Processes were coded for the affected vowels, conditioning environment, and outcome.

As syllable structure complexity increases, so does the likelihood that a language has ongoing processes of vowel reduction (Table 1). The affected vowels, conditioning environments, and outcomes characteristic of these processes also differ according to the syllable structure complexity of the languages in which they occur. A striking result is that languages with HCSS exhibit the highest rates of both vowel reduction in general and the most extreme form of vowel reduction: vowel deletion, typically of a schwa in an unstressed syllable. An analysis of vowel deletion processes in the sample (Table 2) indicates that while these are attested in languages from all four categories, they are significantly more likely to result in tautosyllabic clusters in languages with HCSS than in the other languages.

The findings of this study indicate (1) that syllable structure complexity is correlated not only with segmental properties of languages, but also rates and specific characteristics of active vowel reduction processes, and (2) that active processes of vowel deletion are more likely to create consonant clusters in languages which already have a prevalence of large consonant clusters. These results point to a strong association between phonological, specifically syllable, structure and ongoing phonetic processes. They also suggest that the processes responsible for creating some syllable patterns in the first place can be remarkably persistent in the synchronic phonology of a language.

	Syllable Structure Complexity			
	Simple (N=22)	**Moderately Complex** (N=27)	**Complex** (N=27)	**Highly Complex** (N=24)
Lgs. with V reduction *(%)*	11 *(50%)*	17 *(63%)*	20 *(74%)*	21 *(88%)*
Characteristic affected V	all Vs	high Vs	high Vs	schwa
Characteristic conditioning environment	word/phrase/ utterance position	word position	stress	stress
Characteristic outcome	devoicing	quality reduction	quality reduction	deletion

Table 1. *Rates and characteristics of vowel reduction processes in sample with respect to syllable structure complexity. Trend in first row is significant ($\chi^2(3, N = 100) = 8.34, p < .05$).*

Structural effect of vowel deletion	Syllable Structure Complexity			
	Simple (N=4)	**Moderately Complex** (N=11)	**Complex** (N=8)	**Highly Complex** (N=11)
Tautosyllabic cluster	0	4	3	8
Other outcomes *(e.g., simple onset or coda)*	4	7	5	3

Table 2. *Languages in sample with vowel deletion, and the structural effects of those processes. Trend in Highly Complex category is significant when compared against the other three categories combined (p = .03 in Fisher's Exact test).*

References

[1] Maddieson, Ian. 2013. Syllable Structure. In Dryer, Matthew S. & Haspelmath, Martin (eds.), *The World Atlas of Language Structures Online.* Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at http://wals.info/chapter/12, Accessed on 2016-09-01.)

[2] Maddieson, Ian, Sébastien Flavier, Egidio Marsico, Christophe Coupé, & François Pellegrino. 2013. LAPSyD: Lyon-Albuquerque Phonological Systems Database. In *Proceedings of the Annual Conference of the International Speech Communication Association, Interspeech, Lyon, France.* 3022-3026.

[3] Hill, Jane H. & Ofelia Zepeda. 1992. Derived words in Tohono O'odham. *International Journal of American Linguistics* 58(4): 355-404.

[4] Bagemihl, Bruce. 1991. Syllable structure in Bella Coola. *Linguistic Inquiry* 22(4): 589-646.

[5] Rowicka, Grażyna. 1999. *On ghost vowels - a strict CV approach* (LOT dissertations, 16). The Hague: Holland Academic Graphics.

[6] Chitoran, Ioana & Ayten Babaliyeva. 2007. An acoustic description of high vowel syncope in Lezgian. *International Congress of Phonetic Sciences XVI:* 2153-56.

High boundary tones in spontaneous Southern Swedish.

Anastasia Karlsson[1], David House[2]
[1]*Lund University,* [2]*Royal Institute of Technology, Stockholm*

Our study investigates the relationship between intonation and discourse structure in spontaneous narratives in Southern Swedish. Swedish uses intonation both lexically and postlexically. On the lexical level, two word accents are distinguished, accent 1 and accent 2. There is a systematic variation between dialects in the realization of word accents, described within the dialectal word accent typology (Gårding 1977). Tonal contours in Swedish are composed of lexical tones, prominence tones and boundary tones. There are systematic differences between dialects in the realization (alignment and association) of these tones. Southern Swedish has following composition (adapted from Riad 2006):

	Lexical tone	Prominence tone	Boundary tone
Accent 1		H	L]
Accent 2	L	H	L]

Association of the postlexical tones depends on syntactic and information structure. Postlexical intonation has been best described in Central Swedish, while Southern Swedish has received very little attention in the literature. It has been proposed that Southern Swedish has rather weak tonal cues for focal accentuation showing little difference from basic accentuation patterns (Bruce 2007). Our study is, to our knowledge, the first one that treats Southern Swedish spontaneous speech and investigates accentuation and phrasing in connection with discourse structure. We concentrate on right boundaries and show that their functions and distribution can only be explained in connection to discourse structure. The material consists of 12 spontaneous narratives spoken by six speakers in pairs (three females, three males between 20-30 years of age). The narratives are of two different genres, retelling a movie/picture story and telling about an everyday activity (e.g. cooking). The recordings were analyzed by using three levels, Topic (what the proposition is about) and Comment (information about the Topic), New and Given status of referents, and, finally, prosodic transcription of boundaries and most prominent accents. In the final step, all three levels are compared to determine the connection between them.

High, right-edged boundaries are strikingly frequent in the material. Based on phonetic evidence we differentiate between H% and LH% boundaries. These boundary tones are realized on the last sonorous segment in a word. In monosyllables with prominence tone it gives (L) H LH% and (L) H H% contours. In polysyllabic words H% often results in a high plateau between H and H%.

The high rising terminals could be described as uptalk which is more and more typical for younger people, and we do suppose that we are dealing with uptalk in many cases in our material. However, LH% and H% seem to have different functions, and these functions are connected to discourse structure. H% has a connector function with the following content. LH% occurs when a new Topic is introduced. Old Topic does not usually get LH%, although it may be possible for new referents to receive a LH% tone in a Comment. It is still under investigation whether both boundary tones can have focal function in some contexts (such as new referents in a Topic or Comment).

Southern Swedes have thus to incorporate high boundaries to the prosodic system with an already rather high phonological load of high tones.

References:

Bruce, Gösta. 2007. “Components of a prosodic typology of Swedish intonation”. In *Tones and Tunes*, Volume 1, T. Riad; C. Gussenhoven (eds.). Berlin: Mouton de Gruyter, 113-146.

Gårding, Eva. 1977. *The Scandinavian word accents*. (Travaux de l’institut de linguistique de Lund 11). Lund: CWK Gleerup.

Riad, Tomas. 2006. “Scandinavian accent typology”. In *Sprachtypol*. *Univ*.*Forsch*., STUF, Berlin 59.1: 36–55.

Tuesday, 13:15 - 14:15
Poster Session 3, #9

Tianjin Mandarin Tunes: Production and Perception data

Cong Zhang
Language and Brain Lab, University of Oxford

Alternating pitch accents and boundary tones are common methods of making a yes-no question in non-tonal languages (c.f. English, Greek, etc.); tonal languages, however, are much restricted by lexical tones, including contour shape and register. How does a tonal language speaker ask a yes-no question syntactic marking or even context? Can listeners perceive the questions well? This paper aims to answer these two research questions through investigating the production and perception of the tunes of statements and intonational yes-no questions in Tianjin Mandarin. Tianjin Mandarin is a northern Mandarin dialect which share many common syntactical, phonologically and lexical features with standard Mandarin. Tianjin Mandarin, despite having the same number of lexical tones with standard Mandarin, has a more evenly distributed lexical tone inventory (L, H, LH, HL), which conveniently serves as a good tool for prosody research.

A **production** study was conducted to investigate whether intonational contours would override lexical tone contours and how tone and tune interact. Six native speakers of Tianjin Mandarin (3 male and 3 female) were recorded. Mono syllabic materials were used – three different syllables, each with four lexical tones, were tested for declarative and interrogative tunes. A comprehensive analysis of the data suggests that (**a**) The register is higher in interrogative tunes than in declarative for all tones; (**b**) The pitch range of the intonational yes-no questions is smaller for lexical tones ending with a L tone, but bigger for lexical tones ending with a H tone. This implies that, on the one hand, the Ls in the questions do not fall as sharply as those in the statements; on the other hand, the Hs in the questions rise even higher than those in the statements. This finding is consistent with Chang's (1958, in Ladd 1996) study on Chengdu Chinese questions. (**c**). There is a floating H boundary tone for intonational yes-no questions in Tianjin Mandarin. The literature on standard Mandarin has mixed conclusions. For example, Lin (2004, 2006) assumes boundary tone, while Yuan et al. (2002) assumes none, as the current study does. In summary, intonational yes-no questions in Tianjin Mandarin differs from statements only in terms of **register**, and **a floating H boundary tone** but not in terms of pitch accents. The results coincide with some African languages, which make questions by utilising the register and degree of pitch rise and fall (Rialland, 2007). Tianjin Mandarin also has syntactically marked yes-no questions, in which the utterances end with a question particle. This construction may be a consequence of the fact that intonation alone in a tonal language is not robust enough to facilitate effective communication.

A **perception** experiment was also conducted to investigate whether listeners could identify YNQ from statements by merely using the subtle cues of register change and floating boundary tone. 28 native Tianjin Mandarin speakers (15 male and 13 female) took part in the experiment. They were instructed to perform a forced-choice task by pressing either the 'Q' button or 'S' button when hearing an utterance.

Results: The following table shows the accuracy rate of the identification task. The lexical tones are presented by average accuracy rate. When results from two lexical tones are not statistically significant, the relation is represented with a "=" (Table 1).

The crucial cue for perceiving statements is the L tone at the left boundary. L tone and LH tone both start with a L tone, so they both achieved the highest accuracy. When the initial tone does not help with identification, such as between HL and H tone, the ending H tone interferes with the identification. On the contrary, listeners use the right boundary height for YNQ. The general trend is that if the right boundary is a H tone, then it is easier to identify. Similar to the statement, initial L interferes with the identification too. Since H Tone is a rising H, it is easy to be considered a question. The floating H% added the advantage.

Table 1: Perception experiment results

Sentence type	Gender of stimuli voice	Lexical tones
Statement	Male	LH=L > HL > H
	Female	LH=L=HL > H
Yes-No Question	Male	$H^{H\%}$ > $LH^{H\%}$ =$L^{H\%}$=$HL^{H\%}$
	Female	$HL^{H\%}$ =$H^{H\%}$ > $LH^{H\%}$ > $L^{H\%}$

References

Ladd, D. R. (1996). *Intonational phonology*. Cambridge, MA: Cambridge University Press.

Lin, M. (2004). hanyu yudiao yu shengdiao [Chinese Intonation and Tone]. *Yuyan Wenzi Yingyong*, (3), 57–67. Retrieved from http://en.cnki.com.cn/Article_en/CJFDTOTAL-YYYY200403014.htm

Lin, M. (2006). yiwen he chenshu yuqi yu bianjiediao [Interrogative vs. declarative and the boundary tone in Standard Chinese]. *Zhongguo Yuwen*, (4).

Rialland, A. (2007). Question prosody: an African perspective. In T. Riad & C. Gussenhoven (Eds.), *Tones and tunes* (Vol. 1). Walter de Gruyter.

Yuan, J., Shih, C., & Kochanski, G. P. G. (2002). Comparison of declarative and interrogative intonation in Chinese. In *Speech Prosody 2002, International Conference*.

Tuesday, 13:15 - 14:15
Poster Session 3, #10

Discrimination of German tense and lax vowels in German monolingual and German-Turkish and German-Russian bilingual children

Amrei Walkenhorst
Bielefeld University

The phonetic correlates of the distinctive feature of tenseness in German vowels are duration (long vs. short) and (in all contrasts except /a: a/ and /ɛ: ɛ/) vowel quality (more vs. less peripheral). According to one phonological modelling, the lax vowels appear exclusively in closed syllables ([ˈbe:.tn̩] *to pray* vs. [ˈbɛs.tn̩] *best*). As a consequence, the single internuclear consonant in a word like [ˈbɛtn̩] (*beds*) is interpreted as ambisyllabic, occupying the coda of the first and the onset of the second, reduced syllable [1, 2]. German orthography is in accordance with this phonological modelling, marking the ambisyllabic consonant by doubling the consonant grapheme (<Betten> vs. <beten>) [3]. It is claimed that children with L2 German, compared to children with L1 German, have more problems acquiring this orthographic marker [3]. It is assumed – but has not yet been experimentally tested – that this is caused by lower perceptive discrimination skills due to phonetic/phonological interferences from the L1 [4, 5].

So far, only few studies have investigated vowel perception in young learners of L2 German, but they do not take all of the tenseness contrasts in German into account [e.g. 6]. In the present study, an AX perception task was employed to test perceptive discrimination abilities of all vowel tenseness contrasts except /ɛ: ɛ/. The stimuli were minimal pairs consisting of disyllabic pseudo words with two different consonantal contexts. The participants were monolingual German (G n=22), bilingual German-Turkish (GT, n=37), and German-Russian first-graders (GR, n=16) (6-7 years old) living in the region of Westphalia.

The vowel spaces of Turkish and Russian have less phonemes in comparison to German, tenseness is not a distinctive feature. According to Flege [7] contradictory expectations may be derived from that: (1) lower discrimination performance in GT and GR bilinguals than in German monolinguals due to the identification of a tense and a lax vowel with the nearest Turkish or Russian vowel phoneme or (2) similar or better discrimination performance in those vowel pairs that share vowel space with a Turkish or Russian phoneme due to strict separation of all of the vowel phonemes represented in the bilingual phonological space.

A generalized linear mixed model (GLMM) was fit to the binomial data, with LANGUAGE as fixed factor and participants as random factor. It was revealed that the bilingual children as a group did not discriminate the items less accurately than the monolingual children (G vs. GT: z=.53, p=.60, G vs. GR: z=.48, p=.37, Fig. 1). More detailed inspection into the different vowel categories revealed that the groups showed significantly different results only for the /i: ɪ/ contrast: GT bilinguals discriminated the pair less accurately than the German monolinguals (G vs. GT: z=-2.14, p=.03), while the Russian bilinguals did not differ significantly from the German monolingual group (G vs. GR: z=-1.50, p=.13, Fig. 2). This could be explained by possible L1 interference in auditory perception: Turkish has only one phoneme in the closed fronted unrounded vowel space (/i/), while Russian has two (/i, ɨ/).

The ability to mark vowel tenseness in writing was further investigated by means of a qualitative longitudinal study with a subgroup of the earlier participants (n=12). The results show that some children with high perceptive accuracy rates, among them mono- and bilingual children, mark phonetic details such as allophones and qualitative vowel features (disregarding duration) and seem to have problems to develop a phonological representation of words that the German writing system is based on, including the tenseness contrast. Thus, some childrens' high auditory perceptive skills might impede their acquisition of the German orthographic markers of the vowel contrasts.

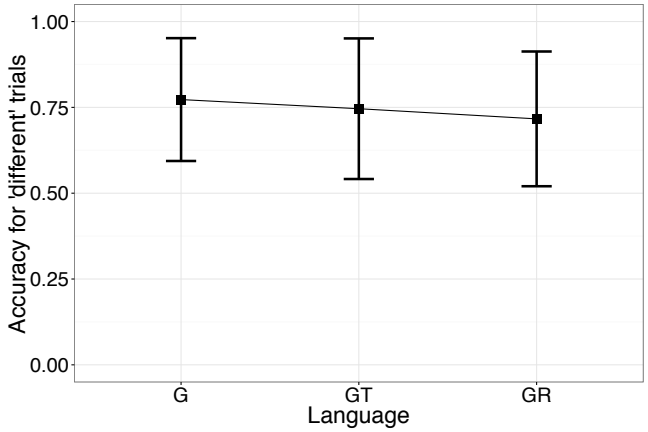


Figure 1. *Accuracy for 'different' trials by language groups*

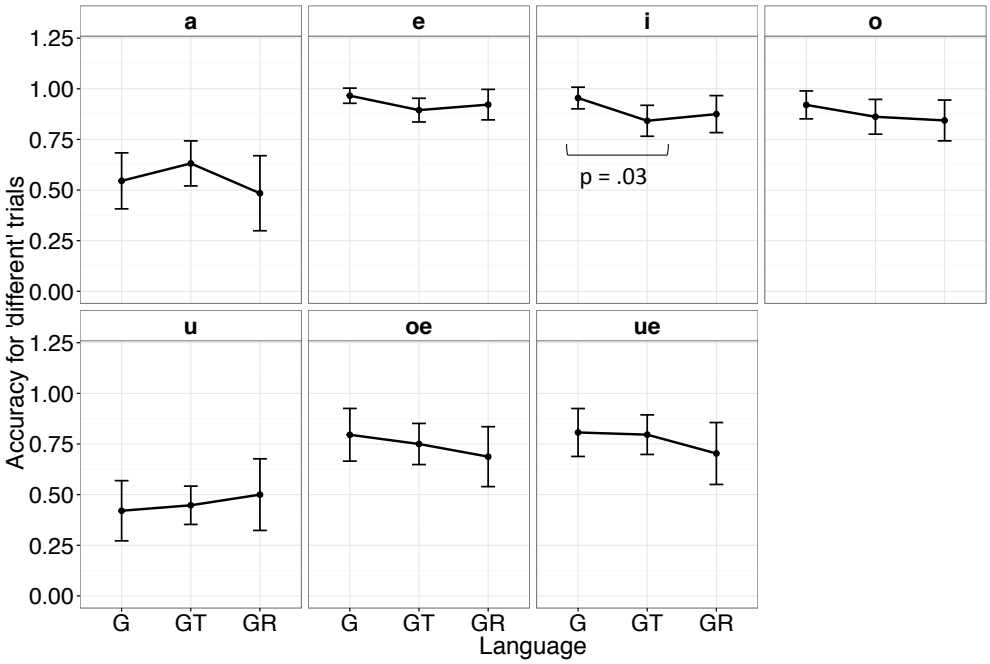


Figure 2. *Accuracy for 'different' trials for the vowel contrasts by language group*

[1] Hall, T. A. 2011. *Phonologie: Eine Einführung*. (2nd ed.). Berlin: De Gruyter. [Phonology. An Introduction].

[2] Wiese, R. (1996). *The phonology of German*. Oxford: Clarendon Press.

[3] Fuhrhop, N. 2009. *Orthografie*. Heidelberg: Winter. [Orthography].

[4] Becker, T. 2011. *Schriftspracherwerb in der Zweitsprache. Eine qualitative Längsschnittstudie.* Baltmannsweiler: Schneider-Verlag Hohengehren [Acquisition of writing in the second language. A qualitative longitudinal study].

[5] Bredel, U. 2013. Silben und Füße im Deutschen und Türkischen – Orthographieerwerb des Deutschen durch türkischsprachige Lerner/innen. In Deppermann, A. (Ed.), *Das Deutsch der Migranten*. Berlin: De Gruyter, 369–390 [Syllables and (metric) feet in German and Turkish – Acquisition of German orthography by turkish-speaking learners].

[6] Darcy, I. & Krüger, F. 2012. Vowel perception and production in Turkish children acquiring L2 German. *Journal of Phonetics*, 40, 568–581.

[7] Flege, J. E. (1995). Second language speech learning: Theory, findings and problems. In W. Strange (Ed.), *Speech perception and linguistic experience. Issues in cross-language research* (pp. 233–277). Baltimore: York Press.

Semi-automatic analyses of vocalic sequences in Romanian

O.E. Niculescu[1], I. Vasilescu[2], B. Vieru[3], L. Lamel[2], M. Adda-Decker[4]
[1]University of Bucharest, [2]LIMSI-Paris Saclay University, [3]Vocapia Research, [4]Paris Sorbonne University

Romanian vowels have been studied both from acoustic [1, 2] and articulatory standpoint [3] using laboratory and large scale data explored with automatic techniques, e.g. forced alignment and statistical modelling. More recently [4, 5], the interest is in the phonemic status of the central vowels /ə/, /ɨ/, or classification of vowels in terms of functional load. In line with such approaches, we propose a study of acoustic and prosodic patterns of vocalic sequences (inter- and intra- word) in Romanian at the interface between phonetics and phonology, with an insight from speech technology.

Romanian has a 7 vowel inventory (/i, ɨ, u, e, ə, o, a/) and two phonologically unary diphthongs (/ea/, /oa/) [6]. We base our analysis on 2 corpora, highlighting the relation between norm vs. usage (the gap between canonical view of vocalic sequences as portrayed by phonological analysis supported by dictionaries and true realization of more than one vowel sequences). The approach is twofold. First, we use a corpus consisting of the prototypical realizations of hiatus vs. diphthong sequences form an acoustic perspective [7] (laboratory speech, logatoms, minimal pairs, manual labelling and parameters extraction), and normative (dictionaries) [8]. We classify hiatus into 4 classes based on height (rising (height V1 > height V2), falling (height V1 < height V2), constant (height V1 = height V2, but V1 V2), identical (height V1 = height V2, V1 =V2)) and localization (anterior, central, posterior), and compare the formant transitions and duration (for each class of hiatus, and between hiatus and diphthong). Second, we extrapolate the findings to a 7 hour broadcast data (141 speakers, recordings from Romanian radio shows, semi-prepared speech and spontaneous debates), forced aligned at phoneme level with the system described by [9], with acoustic parameters extracted following [10]. Within this corpus, we analyse the vocalic sequences both inter- and intra-word in terms of number of occurrences (frequency of vowels in Romanian vs. frequency of vowels in vocalic sequences in the 7h corpus vs. canonical frequency) and duration (we compare both the total duration of the sequence, and the duration of the 2 individual segments). Correlating the results from the two corpora, we observe that, (1) in terms of frequency, the broadcast data confirms the vowel-sequence hierarchy from the normative perspective, in the sense that sequences like /i.e/, /i.a/, /i.i/ are recurrent within the word, as opposed to /e.a/, /a.a/, /ə.a/ which are predominant between words, (2) in terms of duration, for controlled speech, a longer duration and gradual transitions oppose hiatus to diphthongs (in accordance with [11]). On-going analyses comprise in-depth measurements and comparisons of acoustic properties of vocalic sequences as function of speaking condition. More broadly, the results will help for drawing detailed typology of Romanian vowel inventory (both as individual items and through combination patterns with implications at phonetic and phonological level).

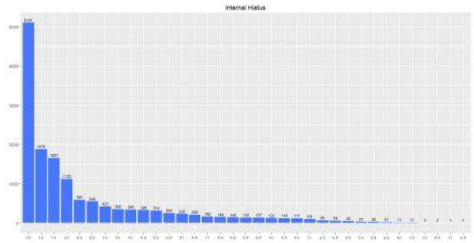


Figure 1.a

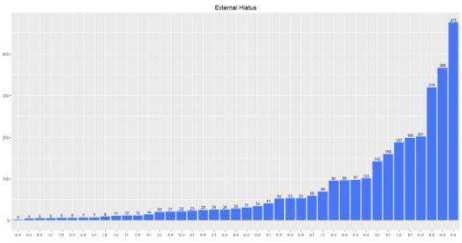


Figure 1.b

Distribution of internal and external hiatus. We observe that the two classes of hiatus are almost in complementary distribution; recurrent sequences for external hiatus (/e.a/, /a.a/) become less predominant for internal hiatus, and vice versa.

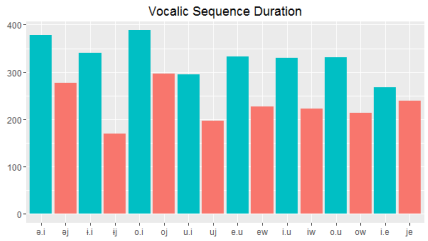


Figure 2. Duration results for hiatus-diphthong minimal pairs

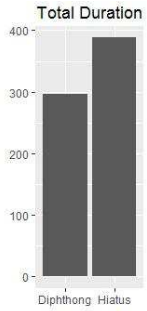


Figure 3. Duration results of hiatus vs diphthong for controlled speech

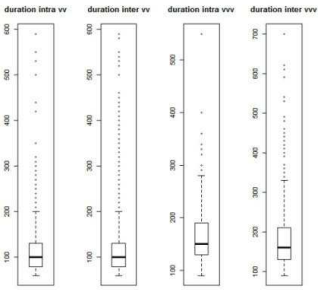


Figure 4. *(Broadcast) patterns of duration independened of the vocalic or semivocalic segment; for VV sequences the duration inter- and intra- word is similar, as opposed to the VVV sequences where the duration intra-word is longer;*

REFERENCES

[1] Spinu, L., I. Vogel and H. Bunnell. 2012. "Palatalization in romanian-accoustic properties and perception". Journal of Phoentics, (40(1)), p. 54–66.

[2] Renwick, M. 2014. *The Phonetics and Phonology of Contrast : The Case of the Romanian Vowel System*. Berlin, Boston : De Gruyter Mouton.

[3] Marin, Ş. 2007. "An articulatory modeling of Romanian diphthongs alternations, ICPhS XVI, p. 453-456.

[4]Vasilescu, I., M. Renwick, C. Dutrey, L. Lamel, and B. Vieru. 2016. "Réalisation phonétique et contraste phonologique marginal : une étude automatique des voyelles du roumain", Actes de la conférence conjointe JEP-TALN-RECITAL 2016, volume 1, p. 597-606.

[5] Renwick, M., I. Vasilescu, C. Dutrey, L. Lamel, and B. Vieru. 2016. "Marginal contrast among Romanian vowels: evidence from ASR and functional load", Proceedings of Interspeech 2016, p. 2433-2437.

[6] Chitoran, I. 2002. *The phonology of Romanian: a Constraint-Based Approach*, Berlin & New York, Mouton de Gruyter.

[7] Niculescu, O. (in preparation) *Hiatul inten şi extern în limba română: o analiză acustică* [Internal and External Hiatus in Romanian: an acoustic study], PhD thesis.

[8] Niculescu, O. 2015, "Resolving Hiatus by Diphthongisation in Romanian", at PaPE 2015 (poster presentation).

[9] Vasilescu, I., B. Vieru, and L. Lamel. 2014. "Exploring pronunciation variants for romanian speech-to-text transcription," *Proceedings of SLTU-2014*, p. 161–168.

[10] Gendrot, Cédric & Martine Adda-Decker. 2005. "Impact of Duration on F1/F2 formant values of oral vowels: an automatic analysis of large broadcast news corpora in French and German". Proceedings of Eurospeech, p. 2453–2456

[11] Aguilar, L. 1999. "Hiatus and diphthong: acoustic cues and speech situation differences", Speech Communication, 28, p. 57–74

Phonotactic and morphonotactic consonant clusters in Standard German German, Standard Austrian German and Standard French

Hannah Leykum, Sylvia Moosmüller
Acoustics Research Institute, Austrian Academy of Sciences

Consonant combinations across morpheme boundaries are termed morphonotactic consonant clusters. In contrast, phonotactic consonant clusters are defined as consonant combinations within a morpheme. As a conclusion of research on especially the processing of phonotactic and morphonotactic consonant clusters [1–3] it is hypothesised that in speech production as well, a differentiation between the two types of clusters is made. Morphonotactic clusters are expected to be less susceptible to reduction and deletion processes as they contain information about the morpheme boundary. However, findings on a possible distinction between phonotactic and morphonotactic consonant clusters in speech productions of speakers of Standard Austrian German (SAG) did not show any differences between the two types of clusters [4–6]. This lack of a distinction between the cluster types could be explained by restrictions emerging from language variety specific timing characteristics which do not allow to reduce consonant clusters regardless of the type of cluster: SAG has to be categorised as a mixed type between quantifying language and word language [7, 8]. In a comparison of word-final consonant clusters in SAG and Standard German German (SGG), the non-significant results could be explained by a high redundancy of the information about the morpheme boundary in both varieties, due to the chosen target words [9].

To further investigate the hypothesis, in the present study word-medial consonant clusters which could occur both as phonotactic and as morphonotactic clusters in German and in French are examined. Word-medial clusters are chosen to reduce the risk of an influence of redundant coding of the information about the morpheme boundary. SGG, SAG and French are investigated to compare three different language types: a word-language, a mixed-type language and a syllable-timed language.

For four different clusters (/sm/, /sl/, /sk/ and /ksp/), word pairs with one word containing the cluster as a phonotactic cluster and the other word with the same cluster across a morpheme boundary are investigated for both languages (e.g. for /sm/: "*Organismus*" vs. "*Missmut*"; "*cosmétique*" vs. "*transmission*"). The target words were embedded in carrier phrases in a postfocal position (read twice by 18 speakers). The intensity and duration values of the target words, the clusters, the individual consonants of each cluster and the surrounding vowels were extracted.

The statistical analyses (mixed-effects models) revealed no influence of the type of cluster on the cluster duration, the duration of the vowel preceding the cluster and most intensity measurements. A significant type-of-cluster*language/variety interaction ($p<0.01$) was found in the absolute duration of the word. A tendency to an interaction between type-of-cluster and gender emerged for the relative duration of the /s/ and for the relative intensity of the /s/.

Even though showing some statistical significant interactions including an influence of the type of cluster, these effects are fairly small, leading to the question whether these differences could be used as a cue by the listeners. In addition, whenever an effect of the type of cluster occurred, the morphonotactic clusters are shorter compared to the phonotactic clusters, which contradicts the hypothesis that morphonotactic clusters are more highlighted in order to make the morpheme boundary salient. Since the investigated languages are characterised by a low morphological richness and since in first language acquisition an effect of the morpheme boundary has been found for morphologically richer languages [10, 11], but not for the morphologically less rich Austrian German [12], an influence of the morpheme boundary on speech production is more likely to occur in languages in which the marking of the morpheme boundary is more important for the understandability. This is worth being investigated in future research.

Dependent variable	Experimental variable	t-value	p-value
Duration of the cluster	Type of cluster	0.787	0.475
Duration of the word	Type-of-cluster*language/variety	5.656	0.009**
Duration of the preceding vowel	Type of cluster	0.221	0.830
Duration of the /s/ in relation to the cluster duration	Gender*type-of-cluster	1.734	0.084 .
Intensity of the cluster	Gender*type-of-cluster	-1.877	0.062 .
Intensity of the cluster in relation to vowel intensity	Type of cluster	-0.027	0.979
Intensity of the word	Type of cluster	-0.121	0.905
Intensity of /s/ in relation to the preceding vowel	Type-of-cluster*gender	-1.808	0.072 .

Table 1: Results of the statistical analyses - concerning effects of the type of cluster

References

[1] Celata C., Korecky-Kröll K., Ricci I. & Dressler W. U. 2015. Phonotactic processing and morpheme boundaries: word-final /Cst/ clusters in German. *Italian Journal of Linguistics*, 27(1), 85–110.

[2] Freiberger E. 2011. Morphonotaktik in der Sprachverarbeitung: Online- und Offline-Tests mit deutschsprachigen Erwachsenen. *Wiener Linguistische Gazette*, 75, 33–52.

[3] Korecky-Kröll K., Dressler W. U., Freiberger E. M., Reinisch E., Mörth K. & Libben G. 2014. Morphonotactic and phonotactic processing in German-speaking adults. *Language Sciences*, 46, 48–58.

[4] Leykum H. & Moosmüller S. 1015. Das (mor-)phonotaktische Konsonantencluster /st/ in wortmedialer und wortfinaler Position in homophonen Wortpaaren. Poster at the *11. Tagung Phonetik und Phonologie im deutschsprachigen Raum* (Marburg).

[5] Leykum H., Moosmüller S. & Dressler W. U. 2015. Homophonous Phonotactic and Morphonotactic Consonant Clusters in Word-Final Position. *Proceedings of the 16th annual conference of International Speech Communication Association (INTERSPEECH)*, 1685–1689.

[6] Leykum H., Moosmüller S. & Dressler W. U. 2015. Word-final (mor-)phonotactic consonant clusters in Standard Austrian German. *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS)*.

[7] Bannert R. 1976. *Mittelbairische Phonologie auf akustischer und perzeptorischer Grundlage*. Lund: Gleerup.

[8] Moosmüller S. & Brandstätter J. 2014. Phonotactic information in the temporal organization of Standard Austrian German and the Viennese dialect. *Language Sciences*, 46, 84–95.

[9] Leykum H. & Moosmüller S. 2016. (Mor-)phonotactic consonant clusters in Standard Austrian German and Standard German German. *Tagungsband der 12. Tagung Phonetik und Phonologie im deutschsprachigen Raum*, 103–106.

[10] Kamandulyté-Merfeldiené L. 2015. Morphonotactics in L1 acquisition of Lithuanian: TD vs. SLI. *Eesti Rakenduslingvistika Ühingu aastaraamat - Estonian Papers in Applied Linguistics*, 11, 95–109.

[11] Zydorowicz P. 2010. Consonant Clusters across Morpheme Boundaries: Polish Morphonotactic Inventory and its Acquisition. *Poznań Studies in Contemporary Linguistics*, 46(4), 565–588.

[12] Freiberger E. 2007. Morphonotaktik im Erstspracherwerb des Deutschen. *Wiener Linguistische Gazette*, 74, 1–23.

The phonology and phonetics of vowel trajectories in English

Renate Raffelsiefen, Anja Geumann
Institut für Deutsche Sprache (IDS), Mannheim, Germany

Vowel trajectories in English (E.) *sigh*, *soy, and sow (pig)* are often classified as "true diphthongs", in contrast to those occurring in *say* and *so*, which have been claimed to pattern with monophthongs ([1], [3], [4], [5], [6], [7], [8]). We argue here that strictly phonological evidence supports the parallel phonemic representations of all of these trajectories in (1a,b) (see next page), where a vowel associated with the nucleus (N) is followed by /i/ or /u/ in the coda (C), thus building a complex rhyme (R). The choice of /i/ and /u/ is motivated by markedness, in accordance with the cross-linguistic affinity between maximal vowel closeness and association with the syllable margin. We further propose that syllabic /i/ and /u/ generally co-associate with the coda in English (cf. (1c)), unlike other monophthongs, which associate only with the nucleus (cf. 1d)).

Strictly phonological evidence for the coda associations in (1) concerns systematic gaps. The neutralization to only the vowels represented as complex rhymes in (1a,b,c) in prevocalic position in English indicates the dominant markedness constraint *NN (No adjacent nuclei). The modifications observed in the loanword adaptation into English in (2) serve to satisfy *NN in that the vowel associating with the first nucleus is either replaced by /i/ or /u/ or is followed by /i/ or /u/, to yield complex rhymes ending in a coda (cf. (1)):

(2) Hebr. *Isr*[a.e]*l* > E. *Isr/*ɛi.ə/*l*, *Isr/*ɪ.ə/l, Span. *p*[a.e]*lla* > E. *p*/ɑi.é/*lla*, French *n*[a.i]ve > E. *n*/ɑi.í/*v*e, French *L*/a.ɔ/*s* > E. *L/*ɛi.ɑ/*s, L/*ʊu/*s*, Russian Царна́ев > E. *Tsarn*/ái.ɛ/*v*

Assuming that "/ɹ/-insertion" in rhotic dialects as in (3a) also serves to satisfy *NN by supplying a morpheme-final coda, the representations in (1) explain the ungrammaticality of /ɹ/-insertion in (3b): no /ɹ/ is inserted when the morpheme ends with a coda already.

(3)a. *the law[ɹ] is, the shah[ɹ] is, the villa[ɹ] is*

b. *the day is, the toe is, the tie is, the cow is, the boy is, the bee is, the zoo is*

The representation of /ɛi/ in (1b) is supported by American E. alternations with /ɛ/ before /ɹ/ in foot-internal position, where high tense vowels are generally banned in English (e.g. /silɪŋ/ <ceiling> : /ʃɪlɪŋ / <shilling>, but */liɹɪk/ (only /lɪɹɪk/ <lyric>)). The alternations can then be explained in terms of a ban on /iɹ/: deletion of /i/ would result in /ɛ/ ("=$_R$" rhymes with):

(4) /pɹɛi/ <pray> vs. /pɹɛɹ/ <prayer> =$_R$ /bɛɹ/ <bear> (cf. /pɹɛiəɹ/ <prayer> 'one who prays')
/ðɛi/ <they> vs. /ðɛɹ/ <they're> =$_R$ /bɛɹ/ <bear> (cf. /ðɛid/ <they'd> =$_R$ /mɛid/ <maid>)

The motivation for positing /ə/ as the initial vowel in /əu/ concerns /t/-allophony in American English. Assuming that vowels in the weak syllable of the foot are restricted to {/i/,/ɪ/,/ə/}, the distribution of flapped versus aspirated /t/ in (5) is explained: foot-internal flapping in (5a) versus foot-initial aspiration in (5b). By contrast, the weak class {/i/,/ɪ/,/ə/,/o/} assumed by those who consider the vowel(s) in *so* a monophthong ([10:153]) seems highly unnatural.

(5)a. (má[ɾ]**ə**u)$_\Sigma$ <motto>, (déi[ɾ]**ə**)$_\Sigma$ <data>, (éi[ɾ]**i**)$_\Sigma$ <eighty>, (náu[ɾ]**ɪ**s)$_\Sigma$ <notice>

b. (lá)$_\Sigma$([t^h]**ɛ**̀i)$_\Sigma$ <latte>, (kʉ́u)$_\Sigma$([t^h]**ʉ**̀u)$_\Sigma$ <kowtow>, (tú)$_\Sigma$([t^h]ù)$_\Sigma$ <tut**u**>, (sǽ)$_\Sigma$([t^h]ə̀ɪ)$_\Sigma$ <satire>

The phonemic representations posited in (1) are supported by measurements of the dynamics (25% - 75% of vowel duration, 6 equidistant steps) of American English vowels/diphthongs based on TIMIT [2], illustrated in Fig. 1 for DR5 Southern Region based on 36 female speakers. Unlike the measurements based only on F2 cited in [1],[8], the trajectories in the F1xF2 plane support the parallel representations in (1a,b,c) versus (1d): complex rhymes are reflected in longer and outward-pointing movements (which are not systematically correlated with duration differences), directed towards the left top corner for the rhymes with /i/ codas, and the right top corner for the rhymes with /u/ codas. By contrast, monophthongs not followed by a coda vowel as in (1d) exhibit shorter downward slopes pointing towards the center, (except for schwa, which exhibits a short outward movement).

(1)a.

R	R	R
/ \	/ \	/ \
N C	N C	N C
| |	| |	| |
/ɑ i/	/ɐ u/	/ɔ i/
(*sigh*)	(*sow*)	(*soy*)

b.

R	R
/ \	/ \
N C	N C
| |	| |
/ɛ i/	/ə u/
(*say*)	(*so*)

c.

R	R
/ \	/ \
N C	N C
\ /	\ /
/i/	/u/
(*bee*)	(*zoo*)

d.

R	R
|	|
N	N
|	|
/ɔ/	/ɑ/
(*law*)	(*shah*)

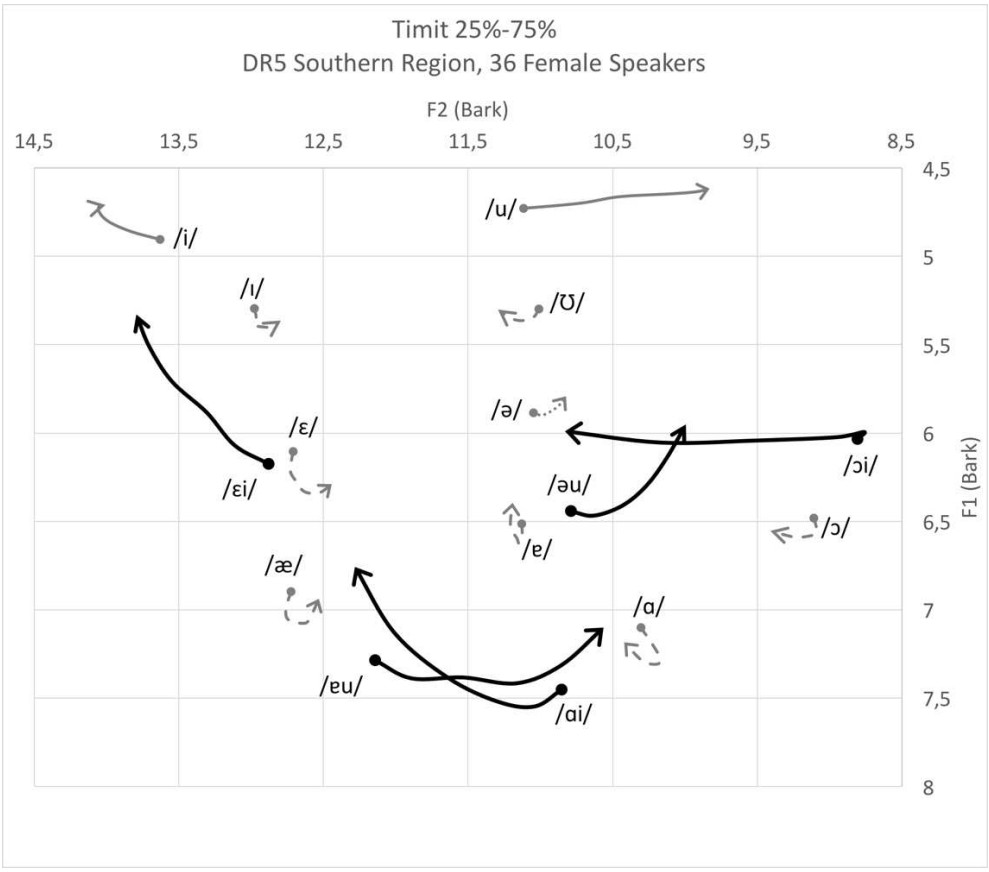


Figure 1. *Formant trajectories for American English.*

[1] Clements, G.N. & Hertz, S.R. 1996. An integrated approach to phonology and phonetics. In Durand, J.: *Current trends in phonology*. Vol. 1: 143-174.

[2] Garofolo, J., et al. 1993. *TIMIT Acoustic-Phonetic Continuous Speech Corpus (LDC93S1)*. Web Download. Philadelphia: Linguistic Data Consortium.

[3] Giegerich, H. J. 1992. *English Phonology. An Introduction.* Cambridge: CUP.

[4] Hammond, M. 1999. *The Phonology of English. A Prosodic Optimality-Theoretic Approach.* Oxford: OUP.

[5] Harris, John 1994. *English Sound Structure.* Oxford, Cambridge: Blackwell.

[6] Kenyon, J. S. & Knott, T. A. 1953. *A Pronouncing Dictionary of American English.* Springfield, Mass., Merriam Company.

[7] Ladefoged, P. 1999. American English. In *Handbook of the International Phonetic Association*. Cambridge: CUP: 41-44.

[8] Lehiste, I. & Peterson, G.E. 1961. Transitions, Glides, and Diphthongs. *Journal of the Acoustical Society of America 33,* 269-277.

[9] Odden, D. 2005. *Introducing phonology.* Cambridge: CUP.

[10] Steriade, D. 2007. Contrast. In de Lacy, P. (ed.). *The Cambridge Handbook of Phonology.* Cambridge: CUP: 139-157.

Tuesday, 13:15 - 14:15
Poster Session 3, #14

Acoustic and perceptual difference between *implicative* and *continuative* intonation in French

Cristel Portes[1], Leonardo Lancia[2]
[1]AMU-CNRS-Laboratoire Parole et Langage, France, [2]Laboratoire de Phonétique et de Phonologie (CNRS, Sorbonne Nouvelle), France

In French intonational phonology, the contrast between the implicative rise-fall (RF) and the continuative rise (R) remains controversial. First, not all inventories of French contours include RF. Delattre (1966) first described this tune and called it "contour d'implication", but some later inventories have two rising-falling contours (Martin 1999, Post 2000, Mertens 2008) while others have only one which does not necessarily corresponds to the implicative RF (Vaissière 1980). On the production side, Post (2000) proposed that the contrast between RF and R relies on an alignment difference of the rise which is delayed on the final syllable for R, and on the addition of a contrastive falling part for RF. The height of both f0 maxima was supposed to be similar. Conversely, Rossi (1999) and Hirst & Di Cristo (1996) conceived RF as derived from R by an increased pitch height. Moreover, while Mertens (2008) argued for a phonological contrast between HB and HB- based on differences in the final height of RF, Ladd (2008: 122) suggests that the fall from the peak "seldom, if ever, reaches the bottom of the range". On the perception side, RF has been claimed to convey an epistemic meaning like obviousness and/or conviction (Portes and Reyle 2014; Delais-Roussarie et al. 2015; Sichel-Bazin, 2015) while R conveys continuation or a polar question.

The aim of this study was therefore to examine the precise phonetic implementation and the differential interpretation of RF versus R. We pursued our aim in three steps. First, we asked to two French phoneticians to identify the instances of RF and R contours in a radio conversation involving five male speakers. Second, we modelled the difference between the two contours via a Wavelet-based Functional Mixed Model (Morris and Carrol, 2006). This technique allowed us to estimate the difference between shapes of f0 trajectories in a mixed model framework, and therefore to implement a complex random effect structure, as required by the unbalanced nature of corpus data. Third, we validated the classification of the two expert listeners with that of 34 naïve French listeners involved in a forced choice identification experiment, using a form-meaning mapping.

The results of the corpus production study showed that: i) two rising-falling contours were found in the data, one corresponding to the implicative RF and the other to a rise-fall with the peak on the penultimate syllable rather than on the final one, ii) some items were judged ambiguous (Amb) between RF and R by the experts, iii)), consistently with Post's (1999) model, the temporal alignment of the rise and the scaling of the falling part but not the scaling of the pitch peak were the relevant cues for the phonetic difference between RF and R (figure 1). The results of the perception experiment (figure 2): i) showed that early aligned RF contours were associated with conviction/obviousness meaning (C/O) while late aligned R contours were associated with continuation meaning (Cont); ii) confirmed that naïve listeners perceived Amb items as ambiguous between implicative RF and continuative R.

This study not only solves a controversy about the relevant contrastive features of the implicative RF compared to the continuative R, but also provides the first experimental validation that naïve listeners associate RF with the expression of conviction and obviousness and R with the indication that the speaker has more to say. Moreover, it shows how, by combining the flexibility of functional data representation with the statistical power of Bayesian analysis, Wavelet-based Functional Mixed Models open the way to a new more data driven approach to the analysis of intonation patterns, even in naturally occurring corpus data.

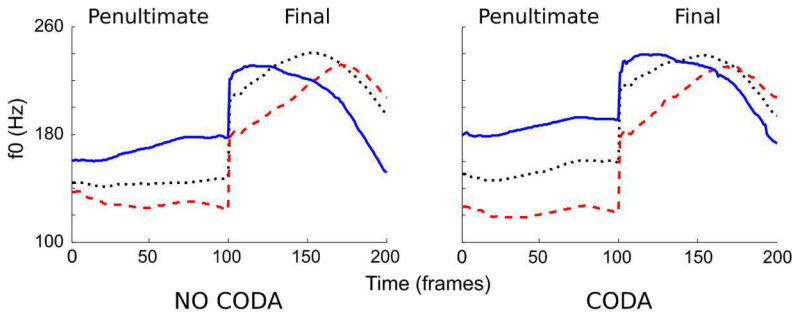


Figure 1: *Estimated trajectories of the f0 curve on the* ***nuclei*** *of the penultimate and the final syllables of RF (solid blue line), Amb (dotted black line) and R (dashed red line). Coda (right) and no coda (left) final syllables were compared but do not differ. A higher f0 on the penultimate signals an earlier beginning of the rise.*

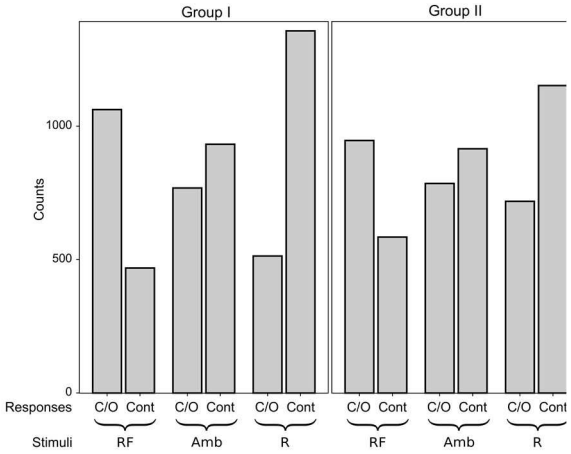


Figure 2: *Counts of responses C/O and Cont over contour type. Left: responses obtained from speakers trained for the hypothesized form-meaning mapping. Right: responses obtained from speakers trained for the reversed form-meaning mapping.*

Delattre, P. 1966. Les dix intonations de base du français. *French Review*, 40, 1-14.

Delais-Roussarie et al. 2015. Intonational Phonology of French: Developing a ToBI system for French. In Sonia Frota & Pilar Prieto, *Intonation in Romance* (Chapter 3). Oxford University Press.

Di Cristo, A., & Hirst, D. 1996. Vers une typologie des unités intonatives du français. In Société française d'acoustique (Ed.), *16émes Journées d'Etude sur la parole*, Avignon, 219-222.

Ladd, R. D. 2008. Intonational Phonology. Second edition: Cambridge University Press.

Martin, P. 1999. Intonation in spontaneous speech in French. In *Proceedings of the14th ICPhS San Francisco*, vol. 1, 17-20.

Mertens, P. 2008. Syntaxe, prosodie et structure informationnelle : une approche prédictive pour l'analyse de l'intonation dans le discours. *Travaux de Linguistique*, 56(1), 87-124.

Morris, J. S., & Carroll, R. J. 2006. Wavelet-based functional mixed models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 68(2), 179-199.

Portes, C., & Reyle, U. 2014. The meaning of French "implication" contour in conversation. In *Proceedings of 7th Speech Prosody 2014*, 20-23 May 2014: Dublin, Ireland, 413-417.

Post, B. 1999. Solving a controversy in the analysis of French rising pitch movements. In *Proceedings of ICPhS1999*, San Francisco, 965-968.

Rossi, M. (1999). L'intonation, le système du français : description et modélisation. Paris: Ophrys.

Sichel-Bazin, R. 2015. *Prosodic systems in contact: Occitan and French*. Unpublished PhD thesis. Universität Osnabrück – Universitat Pompeu Fabra.

Vaissière, J. 1980. La structuration acoustique de la phrase française. In Pacini-Mariotti (Ed.), *Annali Della Scuola Normale Superiore Di Pisa*, Pisa, 530-560.

Tuesday, 14:15 - 14:45
Oral Session 5, #1

Segmental and prosodic effects on perception and production of word-initial clusters

Harim Kwon and Ioana Chitoran
Université Paris Diderot

Previous cross-linguistic research on consonant clusters has typically examined the role of segmental composition in producing and perceiving consonant clusters that do not exist in the speaker-listeners' native language (e.g., [1], [4]). This study aims to investigate the influence of native language on perception and imitation of word-initial consonant clusters, focusing on the effects of the phonetic implementation of clusters and the preferred prosodic structure of the language. Specifically, we ask how native speakers of Georgian spontaneously imitate word-initial clusters produced by a French speaker. Georgian and French differ not only in their phonotactics (Georgian permits more varied onset clusters than French), but also in the phonetic implementation of their onset clusters (Georgian has longer inter-consonant timing lag than French, e.g., [2], [5]). Longer inter-consonant lag in Georgian often results in an epenthetic vowel (a vocalic transition between the two consonants of a cluster) in Georgian native speakers' production ([3]). In addition, prosodically, French has an obligatory word-final (or phrase-final) prominence while Georgian has an initial prominence.

In order to investigate the effects of different inter-consonant timing lags and the preferred prosodic structure, 14 native speakers of Georgian were tested in a spontaneous imitation task. Participants (1) produced CVCV/CCV non-word sequences by reading Georgian script (baseline), and (2) heard and shadowed (immediately repeated what they heard without being told to "imitate") the CVCV/CCV stimuli produced by a French native speaker (shadowing). The auditory stimuli were $C_1V_1C_2V_2$ pseudo-words with 8 different C_1C_2 combinations (bl, gl, pl, kl, sk, sp, ps, pt). V_1 alternated among /a/, /u/, /ø/, and no vowel, while V_2 was always /a/ (e.g., /bala/, /bula/, /bøla/, and /bla/). The "no vowel" stimuli were equivalent to monosyllabic CCVs. During shadowing, each of 32 items (8 C_1C_2 * 4 V_1) was repeated 10 times in differently randomized orders. Participants were not told beforehand that they would hear a foreign language.

Preliminary results (7/14 speakers analyzed) show that the participants' shadowing productions differ from their baseline (reading) productions in several aspects. First, when shadowing French CCV stimuli (or the "no vowel" stimuli which never included an epenthetic vowel), they produced an epenthetic vowel between C_1 and C_2 far less frequently than their baseline productions. 20.2% of CCV tokens in the baseline were produced with an epenthetic vowel, while only 2.3% had an epenthetic vowel in the shadowed CCVs. Moreover, the occurrence of epenthetic vowels as well as their duration decreased with more repetitions (Figure 1, 2). This suggests that the participants, when shadowing French CCV tokens, deviated from their native routines and converged to the French stimuli.

The quality of V_1 in $C_1V_1C_2V_2$ tokens also showed convergence effects: /a/ and /u/ in shadowed CVCV tokens were closer to those in the French stimuli than to those in the Georgian baseline (see Figure 3). On the other hand, /ø/, which does not exist in Georgian, seems to have been assimilated to /u/ although it has higher F2 than intended /u/. 42% of V_1s in shadowed /CøCV/ tokens were labeled as /u/ by two phonetically trained transcribers. Interestingly, 2% of /ø/ were deleted (i.e., /CøCV/ stimuli were shadowed as /CCV/). Although it comprises only a small portion of all the /CøCV/ stimuli, this arguably suggests that Georgian native speakers, whose native language has a longer inter-consonant lag within a cluster, could have perceived "illusory clusters" from French /CøCV/ sequences with word-final prominence. Taken together, we claim that the effects of native language on imitation of word-initial consonant clusters are not limited to the segmental composition of the clusters, but also involve prosodic structure in their phonetic implementation.

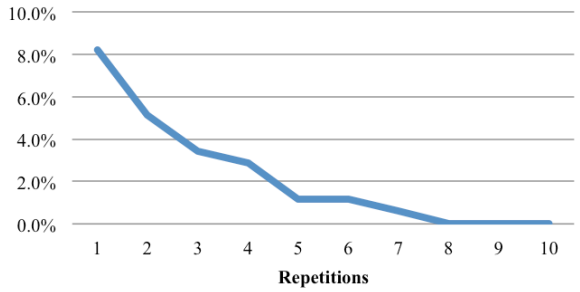


Figure 1. *Occurrence of epenthetic vowels in shadowed French CCV tokens*

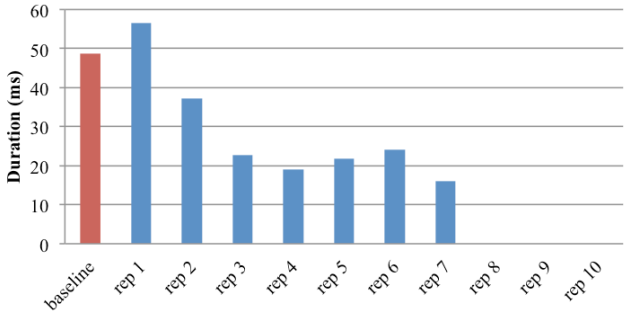


Figure 2. *Epenthetic vowel duration in Georgian (red) and French (blue) CCV tokens*

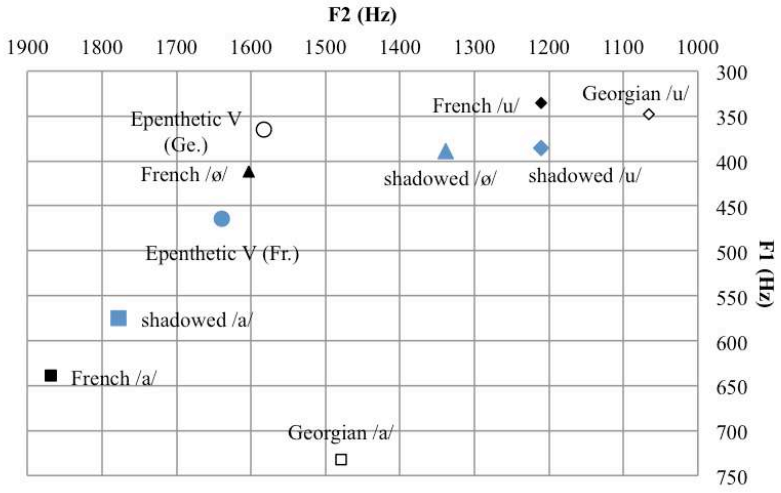


Figure 3. *F1/F2 of produced vowels (black - French stimuli; blank - Georgian baseline; blue - Georgians shadowing French stimuli)*

[1] Berent, I., Steriade, D., Lennertz, T. & Vaknin, V. 2007. What we know about what we have never heard: Evidence from perceptual illusions. *Cognition* 104, 591-630.

[2] Chitoran, I., Goldstein, L. & Byrd, D. 2002. Gestural overlap and recoverability: Articulatory evidence from Georgian. In C. Gussenhoven, T. Rietveld & N. Warner (eds.) *Laboratory Phonology* 7. Berlin: Mouton de Gruyter. 419-447.

[3] Chitoran, I. & Kwon, H. 2016. *Timing lag matters in the perception of Georgian stop sequences by native speakers.* Presented at LabPhon15. Ithaca, NY.

[4] Davidson, L. 2006. Phonology, phonetics, or frequency: Influences on the production of non-native sequences. *Journal of Phonetics* 34, 104-137.

[5] Hoole, P., Bombien, L., Kühnert, B. & Mooshamer, C. 2009. Intrinsic and prosodic effects on articulatory coordination in initial consonant clusters. In G. Fant, H. Fujisaki & J. Shen (eds.) *Frontiers in Phonetics and Speech Science.* The Commercial Press. 275-287.

Vowel insertion in non-native consonant cluster production

I. Sonia d'Apolito, Barbara Gili Fivela
University of Salento – CRIL

This acoustic-articulatory study concerns vowel insertion by Italian advanced learners of French L2 within sibilant clusters across word boundary. Such sequences are very frequent in French, so much so that place assimilations occur [1], and phonotactically marked in Italian (both *per se* and because few consonant clusters are found across word boundary [2]), so that Italian learners may simplify them by inserting a vowel between the two consonants [3;4].

Vowel insertion may be due to: 1) epenthesis (phonological process), with the realization of an articulatory target, or 2) gestural mistiming (phonetic process), when the two consonant gestures are not sufficiently overlapped and, thus, a short vocal tract opening is realized [5;6]. Generally, it is assumed that vowel insertion within an illegal/marked cluster is due to epenthesis [6;7], but experimental studies have shown that it may rather result from gestural mistiming, as learners fail to coordinate accurately the consonant gestures [5;6]. Moreover, speech rate and prosodic structure influence coarticulatory processes [8;9] and thus vowel insertion too.

In this paper, Italian leaners' productions of French sibilant sequences, in two prosodic contexts (weak and strong phrase boundary) and speech rates (normal and fast), were observed to find out if learners repair marked clusters by inserting a vowel and whether the vowel is due to epenthesis or to gestural mistiming. We expect that both a phonological and a phonetic factor play a role: as for the former, 1) learners are expected to insert vowels as a general tendency toward epenthesis in phonotactically marked contexts; as for the latter, 2) timing differences (gestural mistiming) are hypothesized, due to learner attempts to coordinate the consonant gestures; moreover, 3) strong prosodic boundaries may favor insertion and faster speech inhibit insertion.

Heterosyllabic French sibilant sequences, /sʃ; ʃs; sʒ; ʒs; zʃ; zʒ/, were studied acoustically and articulatorily (AG-500), inserted in real words within sentences, in a /aC_1#C_2i/ context. The target words were separated by a weak (phonological phrase) or a strong (intonational) phrase boundary. 3 Italian advanced learners (I1;I2;I3) and 2 French natives (F4;F5) read 7 times the corpus at normal and fast rate. Acoustic and articulatory data were collected simultaneously and analyzed using PRAAT [10] and MATLAB scripts [11] respectively, after a preliminary auditory analysis. All data were segmented and labelled as for the main acoustic (consonant and vowel boundaries) and articulatory events (closing and opening gesture targets, at velocity zero-crossings, on both tongue tip (TT) and dorsum (TD) for consonants and vowels respectively). A series of measurements was performed to assess the coordination of consonant and vowel gestures (Fig. 1).

Auditory and acoustic results show that at normal rate and in both prosodic conditions all speakers insert a vowel. At faster rate I2 and I3 keep inserting a vowel above all in the strong boundary condition, while I1, F4 and F5 assimilate and/or delete the first segment instead (above all in weak boundary condition). On the TD trajectory no specific articulatory target for a vowel was found, for either learners or natives. A further analysis is being carried out on TD to observe the coordination between surrounding vowel gestures and between them and consonant gestures. On the TT trajectory, at normal rate and in both prosodic conditions, the inserted vowel corresponds to an opening gesture between the C_1 and C_2 targets (Fig.2a). At faster rate and in both prosodic conditions, the variability is greater: 1) high degree of gesture overlapping for I1 and the natives F4 and F5 (place assimilation); 2) lower degree of gesture overlapping for I2 and I3, who show 3 articulatory patterns (Fig.2): a) an opening gesture with an acoustic output; b) an opening gesture with no acoustic output; c) no opening gesture and no acoustic output.

Thus, 1) vowel insertion is not the only learner's strategy, but it is frequent and both perceptually and acoustically consistent with an epenthetic (phonological) process; 2) articulatorily it seems to be due to (a phonetic) gestural mistiming; for natives it is far less frequent and found just at normal rate; 3) prosodic structure and speech rate affect vowel insertion as expected.

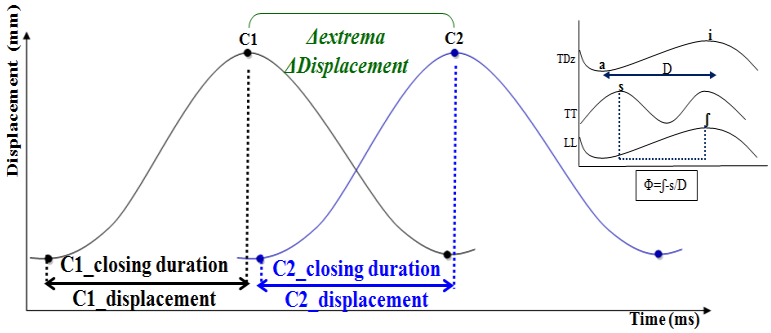


Figure 1. Schema of articulatory measurements: 1) duration (ms) and displacement (mm) of closing gesture for each fricative (*C1/C2 closing_duration; C1/C2 closing_displacement*); 2) latency (ms) and displacement (mm) between the second and the first consonant target (*Δextrema; Δdisplacement*); 3) Relative phasing: ratio between the C2-C1 interval duration and the duration of [a]-[i] articulation (right box)**.**

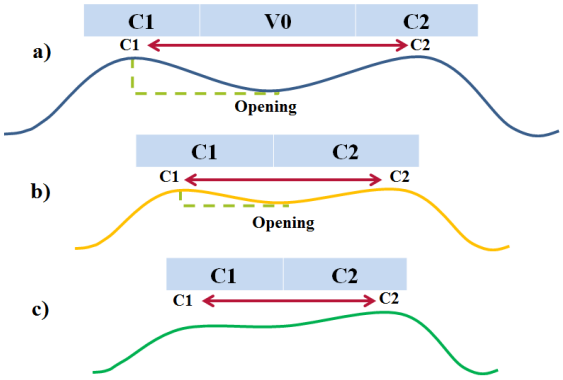


Figure 2. Articulatory patterns. Δextrema (red arrow) shows that a) corresponds to a low degree of overlapping between C1 and C2 with V0 inserted; b) and c) lower degree of overlapping with no V0 inserted.

[1] Niebuhr, O., Lancia, L, Meunier, C., 2008. On place assimilation in French sibilant sequences. *Proceedings of the VII ISSP* (Strasbourg), 221-224.
[2] Muliačić, Ž., 1973., *Fonologia della lingua italiana*, Il Mulino, Bologna.
[3] Renzi, L., 1993. Da dove viene l'articolo il. In Kramer, J.; Plangg, G. A. (Eds.), *Verbum Romanicum: Festschrift für Maria Iliescu*, Hamburg: Helmut Buske, 215-230.
[4] Davidson, L., Stone, M., 2003. Epenthesis versus Gestural mistiming in consonant cluster production. *Proceedings of the West Coast Conference on formal linguistic*, 22 165-178.
[5] Davidson, L., 2006. Phonology, Phonetics, or frequency: Influences on the production of non-native sequences. *Journal of Phonetics*, 34, 104-137.
[6] Davidson, L., 2005. Addressing phonological questions with ultrasound. *Clinical Linguistics & Phonetics*, 19(6/7) 619-633.
[7] Repetti, L., 2012. Consonant-Final Loanwords and Epenthetic Vowels in Italian. *Catalan Journal of Linguistics*, 11, 167-188.
[8] Byrd, D. & Tan, C., C., 1996. Saying consonant clusters quickly. *Journal of Phonetics*, 24, 263-282.
[9] Byrd, D., & Choi, S., 2006. At the juncture of prosody, phonology and phonetics – the interaction of phrasal and syllable structure in shaping the timing of consonant gestures. *Proceedings of the 10th Conference on Laboratory Phonology*.
[10] Boersma, P., Weenink, D., 2010. *Praat: Doing phonetics by computer*, version 5.2.
[11] Sigona, F., Stella, A., Grimaldi, M., Gili Fivela, B., 2015. MAYDAY: A Software for Multimodal Articulatory Data Analys-is. *Atti del X Convegno AISV*, 173-184.

Is the vowel really disappearing in Brazilian Portuguese sandhi? An ultrasound study of vowel reduction.

Francisco Meneses[1], Sarah Johnson[2], Eleonora Albano[1], Ryan Shosted[2]
[1]*Institute of Language Studies, University of Campinas, Brazil*
[2]*Department of Linguistics, University of Illinois at Urbana-Champaign, USA*

This paper examines Brazilian Portuguese (BP) external sandhi, a process previously studied by Bisol (1992, 1996, 2003), Abaurre, Galves & Scarpa (1999), Tenani (2002), and Collischonn (2011). These authors reported three categorical processes involving vowel sandhi at word boundaries in BP: elision (the most common), degemination, and diphthongization. Albano (1999) showed that in some cases where deletion is transcribed, the allegedly elided vowel leaves traces in the acoustic signal. Thus, the pronunciation of, e.g., *cara idoso* 'old guy', has at least three attested acoustic versions: one with a sequence of full vowels, another with a single /i/ vowel resembling *caridoso* 'charitable' and an intermediate version, wherein the values for F1 and F2 just after consonant release show brief evidence of the vowel /a/.

The aim of this study is to investigate, on the basis of acoustic and articulatory data, the occurrence of extreme vowel reduction and/or vowel deletion in BP external sandhi. In order to explore the nature of this process, we designed two experiments. In the first, we recorded three female speakers of BP reading pairs of singleton and cluster Vs in carrier sentences. Duration and formants of the target Vs were measured.

In the second experiment, we designed an automated process for registering relatively large-magnitude deformation in the horizontal (x) and vertical (y) dimensions in sagittal ultrasound images of the tongue (480×600 pixels), based loosely on an algorithm described by Moisik et al. (2010). Three lingual regions of interest were delineated, corresponding to the range of motion of the "front", "mid", and "back" of the tongue (see Figure 1). Velocity functions for each region were logged and visualized in R using customized SSANOVA scripts based on Mielke (2013). For both experiments, sequences of V+V—/a+i/,/a+u/, /i+a+e/—in sandhi positions were compared with their correlates with no sandhi occurrence—/i/,/u/,/ie/—, e.g., *cara idoso* 'old guy' vs. *caridoso* 'charitable' in two speech rate conditions.

The acoustic results show that in normal and even relatively fast speech the vowel deletion in sandhi is less likely to occur, as shown earlier by Albano (1999, 2001). Duration results show a significant difference between V + V and V. These numbers require attention because no statistical differences in fast speech were expected. In addition, the values of F1 and F2 followed exactly the same pattern: both presented significant differences between V+V and V in normal/fast speech.

Likewise, taking into account the vertical displacement, the ultrasound data show that the underlying vowel sequence differs from the singleton vowel, especially in the front and mid tongue regions at the normal speaking rate. In the Figure 2 (right), for instance, SSANOVA shows that /a+i/, in normal rate, is different from /a+i/ fast, but is consistently different from /i/. According to Figure 2 (right), /a+i/ normal is moving is moving faster and higher in mid position. Considering horizontal displacement of the tongue, robust differences were observed in the back region of the tongue. This can be seen in Figure 2 (left), in which the tongue displacement for /i+a+e/, at normal speaking rate is changing from an anterior to a posterior to an anterior position (green spline). This displacement seems to be reduced for /i+a+e/ during fast speech (pink spline). During fast speech, the differences between underlying vowel sequences and singleton vowels appear more reduced in the articulatory results than in the acoustic results. We argue that at fast speech rates, the first (low vowel) gesture is reduced and can be hidden by the high vowel gesture. This extreme reduction can be triggered by increasing the speech rate.

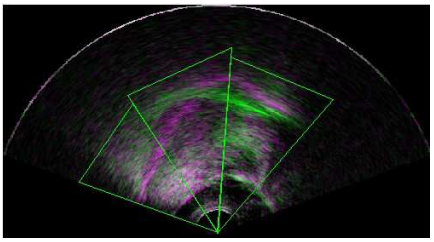


Figure 1. *Triangular regions of interest for the pilot study: "Back", "Mid", and "Front". In this figure, /u/ appears in green, /i/ appears in magenta.*

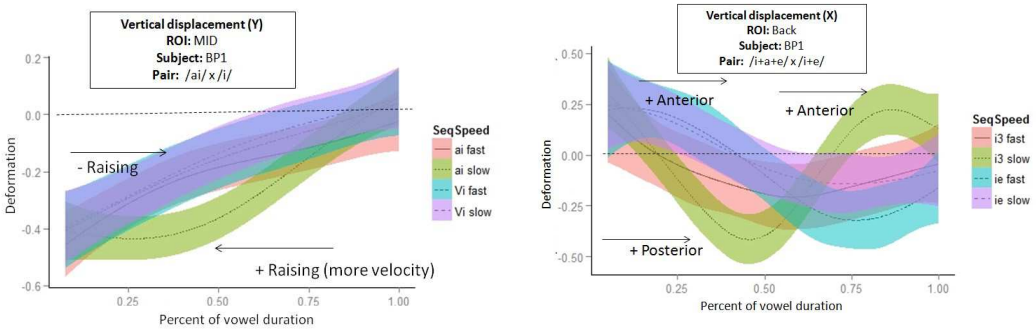


Figure 2. *Examples of SSANOVAs for deformation of two regions of interest (tongue "mid" and tongue "back") for (BP1). Vertical displacement in right figure and horizontal displacement in left figure. Colored bands (95% confidence intervals) correspond to the magnitude of deformation for vowels and vowel sequences at variable speaking rate (Vi=/i/; i3 = /iae/) .*

[1] Bisol, L. 1992. Sândi vocálico externo: degeminação e elisão. *Cadernos de Estudos Linguísticos* 23: 83-101.

[2]_____. 1996. Sândi externo: o processo e a variação. In: ILARI, R. (Ed.). *Gramática do Português Falado IV*. Campinas: Editora da Unicamp.

[3]_____. 2003. Sandhi in Brazilian Portuguese. *Probus* 15(2): 177-200.

[4] Abaurre, M.; Galves, C.; Scarpa, E. 1999. A interface fonologia-sintaxe. Evidências do português brasileiro para uma hipótese top-down na aquisição da linguagem. In: SCARPA, E. (Org). *Estudos de prosódia*. Campinas: Editora da UNICAMP, p. 285-323.

[5] Tenani, L. 2002. Sândi vocálico e estrutura prosódica. *Estudos Linguísticos* 31: 1-4.

[6] Collischonn, G. 2011. A interação entre acento e processos de (re)estruturação silábica: um desafio para a Teoria da Otimidade. *Revista Linguística/Revista do Programa de PósGraduação em Linguística da Universidade Federal do Rio de Janeiro* 7(1): 88-98.

[7] Mielke, J. 2013. tongue_ssanova. Program for R. Downloaded from: http://phon.chass.ncsu.edu/manual/tongue_ssanova.r.

[8] Moisik, S.R.; Lin, H.; Esling, J.H. 2010. An investigation of laryngeal behavior during Mandarin tone production using simultaneous laryngoscopy and laryngeal ultrasound. Proceedings of The 9th Phonetics Conference of China

[9] Albano, E. 1999. O português brasileiro e as controvérsias da fonética atual: pelo aperfeiçoamento da Fonologia Articulatória. *D.E.L.T.A.* [online] 15: 23-50.

[10] Albano, E. 2001. *O gesto e suas bordas*. Campinas: Mercado de Letras/São Paulo: FAPESP.

Speech rate effects on Mexico City Spanish vowel weakening

Meghan Dabkowski
The Ohio State University

Weakening processes that affect the voicing of vowels cross linguistically have been described as occurring more frequently in unstressed high vowels flanked by voiceless consonants (Gordon 1998). Several studies have also found speech rate to contribute to weakening (Dauer 1980 for Modern Greek, Jannedy 1995 for Turkish, among others). The inclusion of the role of speech rate in an analysis of vowel weakening is important because it not only contributes to our understanding of the nature of the process, it also has implications for the model chosen to account for the variation. This paper examines the role of speech rate as well as other factors that predict vowel weakening in the variety of Spanish spoken in Mexico City.

Early characterizations of the phenomenon in this variety reference speech rate as a potential factor conditioning weakening. Canellada and Zamora Vicente (1960) mention that vowels that seem to disappear in fast speech are "restored" in slow, careful speech, and Boyd-Bowman (1952) refers to the phenomenon as occurring in "rapid, nervous speech". Despite these early observations, speech rate effects on vowel weakening have not received a systematic analysis in this variety of Spanish. However, for Cusco Spanish, another variety that exhibits weakening and devoicing, Delforge (2008, 2009) finds that the process is independent of speech rate, although strong claims cannot be made without controlled speech samples.

For this study, the speech of 10 participants was analyzed as part of a larger project investigating vowel weakening in Mexico City Spanish. Among other tasks not analyzed here, each participant completed a reading task which was designed to measure the effects of speech rate on weakening. In this task, participants read a series of 11 sentences three times each: first at a normal rate of speech, then fast, then slow. Sentences were constructed to contain a variety of target vowels in a variety of prosodic and segmental contexts. Examples of the sentences read appear below in (1) and (2). Speeds were self-determined by each participant, following Dauer (1980) and Jannedy (1995). Tokens consisted of any monophthong not adjacent to another vowel, and were acoustically analyzed using Praat (Boersma and Weenink 2016). Measurements were taken for each of the acoustic dimensions that might show reduction: the vowel segment duration, and the voicing duration within that vowel. In addition to speech rate (fast, medium, slow), tokens were also coded for target vowel (/a/, /e/, /i/, /o/, /u/), syllable stress (tonic or atonic), position relative to stress (pre- or post-tonic), and voicing of the segments preceding and following the vowel in question (voiced, voiceless, or pause).

The acoustic analysis indicated a gradient phenomenon: while some tokens showed complete devoicing, others were partially devoiced, produced entirely or partially with weakened voice, or significantly shortened. For the statistical analysis, a generalized linear model was run, collapsing all types of weakened tokens into one category and comparing those to the fully voiced tokens. Results indicate significant effects of position relative to stress, voicing of the following segment, and speech rate. Rapid speech and a following voiceless consonant or pause predicted weakening, while slow speech, post-tonic position, and a following voiced consonant discouraged weakening.

Based on these results, Mexico City Spanish, unlike Cusco Spanish, fits with cross-linguistic trends with regard to the relationship between speech rate and vowel weakening. These results support a dynamic articulatory model of speech, which takes into account the duration and timing of articulatory gestures between adjacent sounds. Within such a model, the observation of more weakening in faster speech is explained by increased gestural overlapping, while less weakening in slower speech is explained by a decrease in the amount of gestural overlapping.

Examples

(1) *Todos mis parientes, menos yo y mis hermanitos, son de Zacatecas.*
All my relatives, except me and my little siblings, are from Zacatecas.

(2) *Me encantan los animales, por lo cual tengo varias mascotas en casa: dos perros, dos gatos, un perico, y cinco peces.*
I love animals, so I have several pets at home: two dogs, two cats, a parakeet, and five fish.

References

Boersma, Paul and Weenink, David 2016. Praat: doing phonetics by computer [Computer program]. Version 6.0.14, retrieved 11 February 2016 from http://www.praat.org/

Boyd-Bowman, Peter. 1952. La pérdida de vocales átonas en la altiplanicie mexicana. *Nueva Revista de Filología Hispánica*, 6, No. 2, 138-140.

Canellada de Zamora, María Josefa & Alonso Zamora Vicente. 1960. Vocales caducas en el español mexicano. *Nueva Revista de Filología Hispánica* 14, 222-241.

Delforge, Ann Marie. 2008a. Unstressed vowel reduction in Andean Spanish. *Selected proceedings of the 3rd Conference on Laboratory Approaches to Spanish Phonology,* ed. by Laura Colantoni and Jeffrey Steele, 107-124. Somerville, MA: Cascadilla Proceedings Project.

Delforge, Ann Marie. 2008b. Gestural Alignment Constraints and Unstressed Vowel Devoicing in Andean Spanish. *Proceedings of the 26th West Coast Conference on Formal Linguistics*, ed. by Charles B. Chang and Hannah J. Haynie, 147-155. Somerville, MA: Cascadilla Proceedings Project.

Delforge, Ann Marie. 2009. *The rise and fall of unstressed vowel reduction in the Spanish of Cusco, Peru: A sociophonetic study*. (Doctoral dissertation, University of California Davis)

Gordon, Matthew. 1998. The phonetics and phonology of non-modal vowels: A cross-linguistic perspective. *Proceedings of the Berkeley Linguistics Society* 24:93-105.

Hundley, James. 1983. *Linguistic variation in Peruvian Spanish: Unstressed vowels and /s/.*

Place of Articulation in Plosives: A Light on the Mapping between Phonetics and Phonology

Daniel McCarthy and Jalal Al-Tamimi
Newcastle University

Phonological contrasts are generally thought of as being discrete, yet the acoustic dimensions that underpin such contrasts are gradient. Given this categoricity of the phonological side of the coin, some researchers (e.g. [2]) have contended that the acoustic attributes that differentiate phonemes should show categoricity also, i.e. the acoustic attributes of a given phoneme should show some invariance across all phonetic contexts. But such invariance has remained elusive. As a result, others have argued that instead of trying to find absolute invariance, we should instead prioritize a combination of acoustic features if it yields sufficient discriminability between phonemes [3]. This is sometimes referred to as relational invariance.

The present study examines these alternative claims on the mapping between phonology and acoustic phonetics by investigating the acoustics of plosives' places of articulation. How best to identify the place of articulation of such consonants has attracted much attention, yet their identification remains far from fool-proof (e.g. [5]'s methods misclassified the place of English plosives at a rate of 8%, [1] at a rate of 10%). Although there are two main cues for identifying a plosive's place – its release burst and its formant transitions – the formant transitions have tended to be neglected by automatic speech recognition (see e.g. [1], p. 838). Linked to this, ASR has normally found the burst to be a better cue than the transitions. But as one of these researchers concedes, '[…] it is possible that the inferiority of the formant-related attributes might be the result of our choice of acoustic measurements used to capture the formant information' ([5], p. 161).

The present study (using discriminant analysis and leave-one-out cross-validation) compares the performance of some of the best-known methods for identifying plosive's place of articulation [4, 5] with some new ones devised by the first author. These latter compare the formant frequency of the vowel onsets and offsets with the frequencies of the vowels' midpoints and/or the peak frequency of the burst. The data consist of a manually annotated corpus of sentences read by 20 speakers of British English (10 male, 10 female), and contains a total of ca. 5,400 plosive tokens, which occur in a wide range of phonetic contexts (e.g. in stressed and unstressed syllables, before front, back and central vowels, and as part of consonant clusters).

Although provisional results reveal that most of the new formulae do not perform better than existing ones, a combination of some of the old and new ones yields a classification accuracy for the three places of articulation of 94%. One interesting result is that the frequency of the burst's peak (= component of greatest intensity) correctly distinguishes velars from alveolars at a rate of 84%, but this improves to 89% when the peak frequency is subtracted from the frequency of the following vowel's midpoint F2. This improvement suggests that relational invariance is a more promising approach than seeking absolute invariance: cues to phonological contrasts become more robust when they are calibrated with reference to the following phonetic context.

References

[1] Ali, A.M.A., Van der Spiegel, J., and Mueller, P. 2001. Robust Classification of Consonants using auditory-based speech processing. *International Conference on Acoustic, Speech, and Signal Processing*, Salt Lake City.

[2] Blumstein, S.E. and Stevens, K.N. 1979. Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants. *Journal of the Acoustical Society of America*, 66(4), 1001-1017.

[3] Lindblom, B. 1990. Explaining phonetic variation: A sketch of the H&H theory. In Hardcastle, W. and Marchal, A. (Eds.), *Speech Production and Speech Modeling*, 403-439. Kluwer, Dordrecht.

[4] Stevens, K.N., Manuel, S.Y., and Matthies, M.L. 1999. Revisiting Place of Articulation Measures for Stop Consonants: Implications for Models of Consonant Production. *Proceedings of the 14th International Congress of Phonetic Sciences*, Volume 2, 1117-1120.

[5] Suchato, A. (2004) *Classification of Stop Consonant Place of Articulation*. PhD Thesis, Massachusetts Institute of Technology.

The development of prosodic and gesture cues to focus in French pre-schoolers

Núria Esteve-Gibert[1], Hélène Lœvenbruck[2], Marion Dohen[3], Thierry Legou[1] and Mariapaola D'Imperio[1]

[1]*Aix Marseille Univ, CNRS, LPL, Aix-en-Provence, France,* [2]*Univ. Grenoble Alpes, CNRS, LPNC UMR 2105, Grenoble, France,* [3]*Univ. Grenoble Alpes, CNRS, GIPSA-lab, Grenoble, France*

This study investigates how pre-school French children make use of prosodic and gesture cues to produce sentences in broad focus and narrow (contrastive and corrective) focus conditions. Children's ability to use prosody for signalling the information status of discourse elements is not mastered until late in development. English-, German-, or Dutch-speaking children, for instance, do not use adult-like prosodic cues to focus before the age of 8-10 years [1]–[3]. Yet, recent evidence shows that these children use phrasing strategies (pauses) to signal new referents in the discourse when they are only 2 years of age [4]. As for gestures, children's use of gestures to focus discourse elements has not investigated yet. However, we know that adults can use head nods and eyebrow movements to mark new elements in the discourse [5]–[7], these gestures co-occurring with prosodic cues to focus [8]. Our study aims examines how French children develop their control of prosodic cues for focus-marking, and whether and how they integrate gestures with prosody in doing so.

Forty-four French 4- and 5-year old preschoolers (mean age: 4;11; age range: 3;11-5;8) participated in a game in which they had to tell to a virtual character which object had to be taken out of a bag in order to trigger a playful activity (Figure 1). Spontaneous sentences like *Prends le bonnet violet* 'Take the purple hat' were elicited. We manipulated the number and type of objects inside the bag to elicit 3 conditions (broad focus; narrow focus: contrastive; narrow focus: corrective) and 2 target focus positions (focused-noun; focused-adjective). A control group of 18 French-speaking adults were also tested using the same procedure. Children were audio-visually recorded and their linguistic and hearing abilities were assessed. Adults were also audio-visually recorded and additional gesture measures were obtained using EMA sensors attached to their face. We analyzed prosodic (pauses; F0max; F0min; duration of words, syllables, and segments; pitch contour) and gestural cues (gesture type; gesture-speech alignment).

Preliminary results reveal that younger children (4 year olds) tended to produce pitch and duration cues of narrow focus (contrastive and corrective) on the last word of the sentence (the adjective) in both the focused-noun and focused-adjective conditions, along with pauses at times produced before the focused element. Interestingly, the gesture cues of narrow focus tended to be more flexible position-wise, since younger children could produce head nods and eyebrow movements co-occurring with the target focused word (be it the pre-final noun or the sentence final adjective). Older children (5 year olds) seemed to be better at using prosodic and gesture cues for narrow focus, marking focused elements with duration (lengthening of segments, syllables and words), pitch cues (F0 max and intonation contour), and co-speech head nods and eyebrow movements, irrespectively of the position of the focused element.

If preliminary observations are confirmed, our findings will confirm that adult-like development of focus occurs late in development [1], [2], although young children manage to mark focus by exploit other prosodic cues [4]. This will also be the first study examining gesture cues next to prosodic cues, and suggesting that young children can do focus-marking if information highlighting as a multimodal phenomenon. All in all, we will add on previous research on the dynamic emergence of prosodic and gestural abilities for meaning purposes in children.

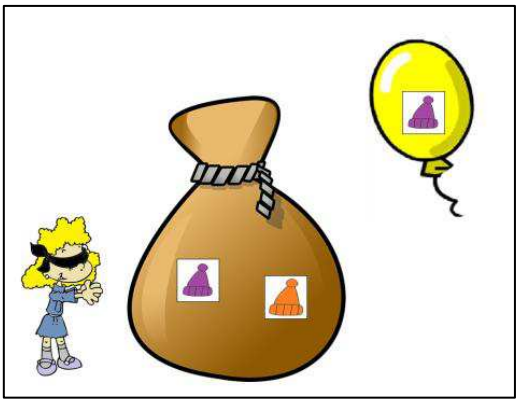


Figure 1. *Example of a visual scene during a trial in the contrastive focused-adjective condition (expected sentence:* Prends le bonnet VIOLET *'Take the PURPLE hat'). On the top right corner, the target object is presented together with the target activity (making the balloon explode).*

[1] K. Ito, "Children's pragmatic use of prosodic prominence," in *Pragmatic Development in Frist Language Acquisition*, 2014.

[2] A. Chen, "Tuning information packaging: intonational realization of topic and focus in child Dutch," *J. Child Lang.*, vol. 38, pp. 1055–1083, 2011.

[3] L. E. De Ruiter, "The prosodic marking of topical referents in the German 'vorfeld' by children and adults," *Linguist. Rev.*, vol. 26, pp. 329–354, 2009.

[4] A. S. H. Romoren and A. Chen, "Quiet is the New Loud: Pausing and Focus in Child and Adult Dutch," *Lang. Speech*, vol. 58, pp. 8–23, 2015.

[5] J. Kim, E. Cvejic, and C. Davis, "Tracking eyebrows and head gestures associated with spoken prosody," *Speech Commun.*, vol. 57, pp. 317–330, 2014.

[6] B. House, D., Beskow, J., Granström, "Timing and interaction of visual cues for prominence in audiovisual speech perception," in *Proceedings of Eurospeech*, pp. 387–390, 2001.

[7] E. Krahmer and M. Swerts, "The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception," *J. Mem. Lang.*, vol. 57, num. 3, pp. 396–414, 2007.

[8] N. Esteve-Gibert, J. Borràs-Comes, M. Swerts, and P. Prieto, "Head gesture timing is constrained by prosodic structure," in *Proceedings of the Speech Prosody Conference*, Dublin (Ireland), 2014.

Tuesday, 16:00 - 17:00
Poster Session 4, #4

Lexical Retrieval and Effects of Homophones

Chelsea Sanker
Brown University

Introduction: This experiment investigated how the presence of homophones within a lexicon or a stimulus set influences response times in listening tasks and how those effects can shed light on lexical storage and processing. Homophones can exhibit patterns in phonetic differences,[2] which listeners may be aware of to some degree. Listeners can identify non-contrastive differences, though more slowly than phonological contrasts.[1]

Response times are influenced by whether homophone-homophone pairs are present within the stimuli, where differences lie within 'different' pairs, and whether a particular word has a homophone, as well as by the type of pair. These patterns suggest that a lexical activation model needs to account for how homophones fit into phonological neighborhoods. They also demonstrate the importance of listening context in weighting activation.

Methods: Same-different task. 15 participants, native English speakers. Stimuli from sentences with similar phonological environments. Pairs of English words, with each word in the pair from a different speaker: (a) homophone-homophone pairs (e.g. *pale-pail*); (b) different pairs in which one of the words has a homophone (e.g. *pale-fail*); (c) different pairs in which neither word has a homophone (e.g. *fail-feel*); (d) same pairs for a word with a homophone (e.g. *pale-pale*); (e) same pairs for a word with no homophone (e.g. *fail-fail*). Within a block, differences were always in the same position (onset, nucleus, or coda). Half of the blocks contained no hph-hph pairs; the other half contained all five pair types.

Results and Discussion: Context influences expectations (cf. [6]). The mean response time was slower in blocks with hph-hph pairs, even if those pairs are excluded in calculations (1900 ms vs. 1690 ms, $p < 0.0001$). The effect is strongest in nucleus-diff. blocks. Though hph-hph pairs were identified as 'same' as consistently as same pairs (92%), they may have slowed down decisions by increasing attention to sub-phonemic differences.

Responses in nucleus-diff. blocks were significantly slower than in coda-diff. blocks (1940 ms vs. 1740 ms, $p < 0.0001$), which in turn were significantly slower than in onset-diff. blocks (1740 ms vs. 1650 ms, $p = 0.00033$); see Figure 1. The speed of onset decisions can be attributed to the position of onsets. The slow responses for nucleus decisions suggest greater uncertainty about vowel contrasts, perhaps due to greater variability or category overlap.

Responses to hph-hph pairs were slower than responses to the other pair types (1990 ms vs. 1770 ms, $p < 0.0001$), suggesting an effect of the sub-phonemic differences in production due to each homophone's frequency and other distinctions. The slowness of responses to hph-hph pairs was strongest in nucleus-diff. blocks (2330 ms vs. 1900 ms, $p = 0.00033$); it was not significant in other block types. Listeners made faster decisions in 'same' trials than 'different' trials (1740 ms vs. 1800 ms, $p < 0.0044$), consistent with previous work (e.g. [5]).

Responses were slower for same pairs of words for which a homophone exists, i.e. pair type (d) vs. (e): 1790 ms vs. 1710 ms ($p = 0.034$); the difference was strongest in nucleus-diff. blocks. Hesitance to identify these as 'same' may reflect listeners' uncertainty about whether homophones are identical or just close phonological neighbors; cf. slower responses for words with high neighborhood density.[3,4]

Conclusions: Existing neighborhood activation models of lexical recognition do not account for homophones [e.g. 4,6]. The differences in response times for words which do and do not have homophones and the effect of the presence of hph-hph pairs within a block indicate the importance of including homophones in processing models. The particularly large effect of homophones in nucleus-diff. blocks suggests that listeners look to nuclei for differences between homophones. Differences between block types demonstrate the effect of expectations on processing, and moreover indicate that contrasts do not all behave similarly.

	Df	Sum Sq	Mean Sq	F value	P value
PairType (same, different, etc.)	4	50.0	12.5	11.9	1.21e-09***
BlockType (onset, nucl., coda)	2	130	64.8	61.7	< 2.2e-16 ***
HomType (homographs, non-hgs)	1	0.0	0.012	0.0112	0.92
StimSet (with/without hph-hph pairs)	1	94.4	94.4	89.9	< 2.2e-16 ***
Participant	14	1545	110.4	105.0	< 2.2e-16 ***
PairType:BlockType	8	31.6	3.95	3.75	0.00021 ***
PairType:HomType	2	1.8	0.898	0.854	0.426
BlockType:HomType	2	2.8	1.42	1.35	0.26
PairType:StimSet	4	4.0	0.998	0.949	0.434
BlockType:StimSet	2	8.9	4.47	4.26	0.0142*
HomType:StimSet	1	2.8	2.79	2.65	0.104
Residuals	9303	9780	1.05		

Table 1. *ANOVA of response times*.

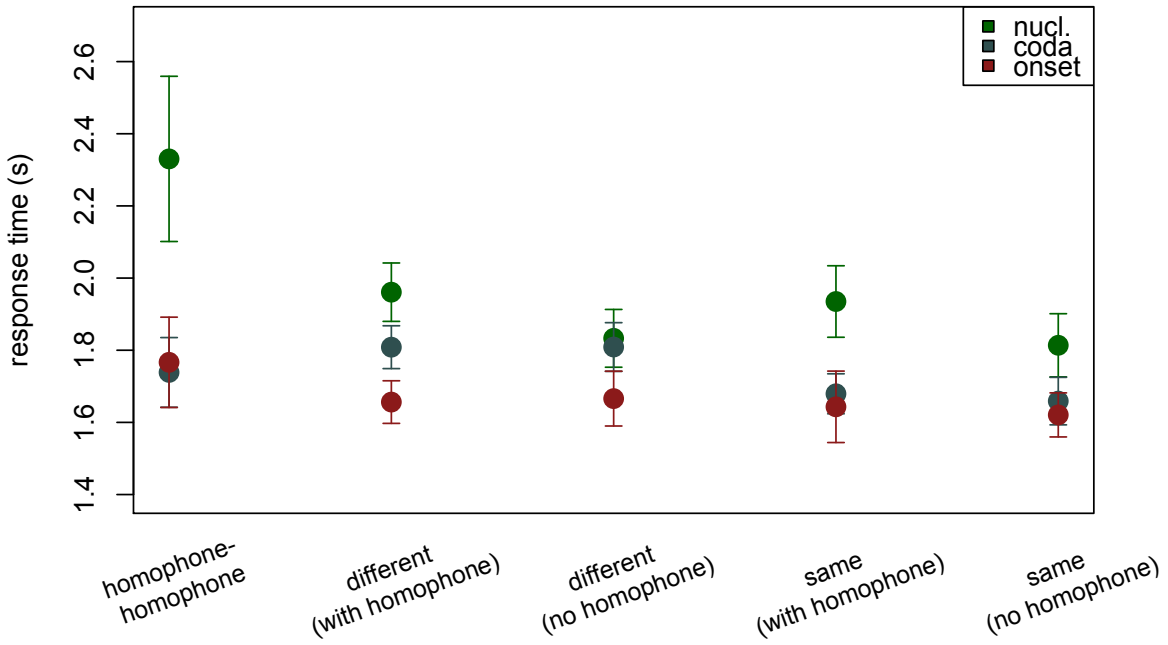


Figure 1. *Response times by pair type and block type*.

[1] Babel, M., & Johnson, K. 2010. Accessing psycho-acoustic perception and language-specific perception with speech sounds. *Laboratory Phonology*, 1, 179–205.

[2] Gahl, S. 2008. *Time* and *thyme* are not homophones: The effect of lemma frequency on word durations in spontaneous speech. *Language*, 84(3), 474–498.

[3] Luce, P., & Large, N. 2001. Phonotactics, density, and entropy in spoken word recognition. *Language and Cognitive Processes*, 16(5-6), 565–581.

[4] Luce, P, & Pisoni, D. 1998. Recognizing spoken words : The neighborhood activation model. *Ear and Hearing*, 19(1), 1–36.

[5] Reed, C. 1975. Reaction times for a same-different discrimination of vowel-consonant syllables. *Perception & Psychophysics*, 18(2), 65–70.

[6] Vitevitch, M., & Luce, P. 1999. Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language*, 40, 374–408.

The detection of semantic incongruity on post-focal constituents in Italian

Jacopo Torregrossa[1], Diana Dimitrova[2], Martine Grice[2]
[1]*University of Hamburg*, [2]*University of Cologne*

It is claimed that contrary to Germanic languages, Italian does not allow for deaccentuation of given information, especially if it is realized after the focused word, that is as post-focus (see (1)). This claim is supported by the prosodic analysis of Italian right-dislocated, post-focal constituents (Bocci & Avesani 2011), as well as by the encoding of givenness in noun phrases by advanced Italian speakers of Dutch (Swerts et al. 2002) and German (Avesani et al. 2015). These studies show that Italian speakers accent given information, albeit using an accent with little or no pitch excursion. Several studies on Germanic languages have argued that constituents bearing a focal accent are processed more deeply than ones that are deaccented (e.g., Sanford & Graesser 2006) and that listeners allocate more attention to anomalous words when they bear a focal accent (Wang et al. 2009). The present study investigates whether the accentual marking of focus type in Italian affects the detection and processing of semantic incongruity in single sentences.

Per comprare la carne, …
(1) bisogna dare I SOLDI *al macellaio/all'infermiere* nel negozio. (post-focus)
(2) bisogna dare i soldi AL MACELLAIO/ALL'INFERMIERE nel negozio. (narrow focus)
(3) bisogna dare i soldi al macellaio/all'infermiere nel negozio. (broad focus)
'To buy meat, we should give the money to the butcher/the nurse in the shop.'

In a behavioral study, 28 speakers of a central Italian variety (Macerata, Marche; age range 19-59; mean 34 years) listened to pre-recorded sentences where two factors were manipulated: semantic congruity (congruous vs. incongruous information) and focus type (broad focus, narrow focus and post-focus) (6 conditions x 20 items, see Table 1). The 120 target sentences were combined with 80 fillers and recorded by an Italian speaker of the Macerata variety. We tested if listeners detect semantic incongruity (e.g. the nurse instead of the butcher in examples 1-3) more easily when it occurs in broad and narrow focus (2, 3) vs. post-focus (1). Participants listened to single sentences and judged whether the sentence was meaningful or not by pressing a yes-no button at the sentence end. Accuracy and reaction times (RTs) were recorded.

We performed a 2x3 repeated ANOVA with RTs as a dependent variable and semantic congruity and focus type as independent ones. The ANOVA analysis revealed a significant effect of *semantic congruity* ($F(1,17) = 9.78$, $p = .001$), showing that responses in incongruous contexts were overall faster than responses in congruous contexts. There was also a significant interaction of *focus type* x *semantic congruity* ($F(2,34) = 7.63$, $p = 0.01$). As Figure 1 shows, participants detected semantic incongruity faster when the incongruous word was in broad and narrow focus as compared to post-focus. The ANOVA analysis of accuracy scores did not show any effects of *semantic congruity*, *focus type*, nor any *focus type* x *semantic congruity* interaction (see Table 2).

The results of the present study are consistent with findings for Germanic languages in showing that focal pitch accents (narrow and broad) affect the processing of the words they are on, speeding up the detection of incongruous stimuli. By contrast, incongruous post-focal constituents take longer to process, as indicated by the slower reaction times. Thus, for Italian, the processing of focused constituents differs from that of post-focal constituents, despite the fact that post-focal constituents are not generally considered to be deaccented.

	Condition	Example
1	Broad F, congruent	Per comprare la carne, bisogna dare i soldi *al macellaio* nel negozio. (To buy the meat, we should give the money to the butcher in the shop)
2	Broad F, incongruent	Per comprare la carne, bisogna dare i soldi *all'infermiere* nel negozio. (To buy the meat, we should give the money to the male nurse in the shop)
3	Narrow F, congruent	Per comprare la carne, bisogna dare i soldi *AL MACELLAIO* nel negozio.
4	Narrow F, incongruent	Per comprare la carne, bisogna dare i soldi all'*INFERMIERE* nel negozio.
5	Post-F, congruent	Per comprare la carne, bisogna dare I SOLDI *al macellaio* nel negozio.
6	Post-F, incongruent	Per comprare la carne, bisogna dare I SOLDI *all'infermiere* nel negozio.

Table 1: *Examples of stimuli for each condition (F=focus).*

	Congruent	Incongruent
Broad F	74.6	76.8
Narrow F	76.1	74.3
Post-F	76.1	73.4

Table 2: *Accuracy scores (in %) in semantically congruous and incongruous contexts across the three focus type conditions (broad, narrow and post-focus) (F=focus).*

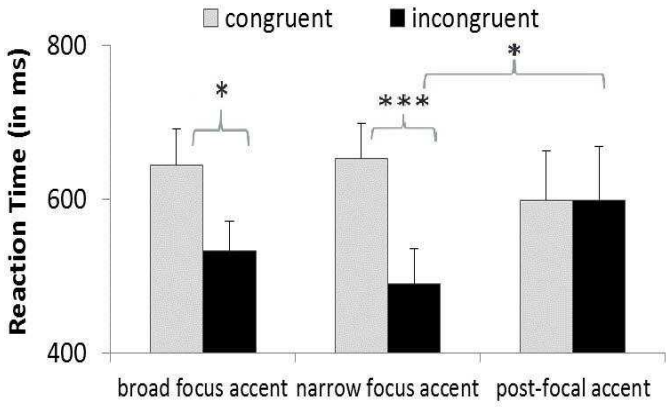


Figure 1: *Reaction times in semantically congruous and incongruous contexts across the three focus type conditions (broad, narrow and post-focus).*

References

Avesani, C., Bocci, G., Vayra, M. & Zappoli, A. 2015. Prosody and information status in Italian and German L2 intonation. In Chini, M. (ed.), *Il parlato in Italiano L2: Aspetti pragmatici e prosodici*. Franco Angeli: Milano.

Bocci, G. & Avesani, C. 2011. Phrasal prominences do not meet pitch movement: postfocal phrasal heads in Italian. *Interspeech* 2011: 1357-1360.

Sanford, A. & Graesser, A.C. 2006. Shallow processing and underspecification. *Discourse Processes* 42(2): 99-108.

Swerts, M., Krahmer, E., Avesani, C. 2002. Prosodic marking of information status in Dutch and Italian: A comparative analysis. *Journal of Phonetics* 30(4): 629-654.

Wang, L., Hagoort, P., Yang, Y. 2009. Semantic illusion depends on information structure: ERP evidence. *Brain Research* 1282(28): 50-56.

Tuesday, 16:00 - 17:00
Poster Session 4, #6

Three Dimensions of Sentence Prosody and their (Non-)Interactions

Michael Wagner & Michael McAuliffe
McGill University

Prosody simultaneously encodes different kinds of information, including the type of speech act of an utterance (e.g., through declarative vs. interrogative intonation), the location of semantic focus (through its effects on prosodic prominence), and syntactic constituent structure (through its effect on prosodic phrasing). The syntactic/semantic functional dimensions (speech act, focus, constituency) are orthogonal to each other, but to which extent their prosodic correlates (tune, prominence, phrasing) are remains controversial.

Two broad types of models about how the different dimensions relate to each other can be distinguished: 'Overlay' models (Öhmann, 1967; Fujisaki, 1981; Xu, 2005) assume that prominence, phrasing, are independent functions that affect acoustic parameters. Most phonological models starting with Pierrehumbert (1980), on the other hand, hold that the effects of these functions are mediated by an auto-segmental phonological representation (the 'AM-Model', cf. Ladd 2008) and predict specific **interactions**.

This study presents a 'bottom up' approach to establish which interactions between the dimensions are in fact attested. A factorial production experiment involved coordinate structures crossed the three functions: speech act, syntactic constituent structure, focus (27 speakers, tested on 4 different sets of items, varying between 16 conditions). Each utterance involved a coordinate structure that was either left ([AB]C) or right branching (A[BC]), see example (1)). Perhaps the strongest test between the models involves the contributions of cues to phrasing (reflecting syntactic constituent structure) by way. According to ToBI and other AM-models, there should not be any pitch accents in the post-focal domain. Beckman (1996) argues that "there must be at least one pitch accent somewhere in every (prosodic) phrase", hence in the unaccented, postfocal domain phrasing distinctions should be neutralized. Similarly, rising polar questions involve L* accents rather than H* accents, hence the scaling of accents should work differently (if phrasing is cued at all with pitch in polar questions, a fact that has not been established). We looked both for effects on the pitch scaling (Ladd, 1988, cf.) in the postfocal domain and on duration (cf. Norcliffe and Jaeger, 2005).

Results. Although raw pitch does not look like a consistent cue to phrasing (Fig. 1), with very different patterns across tunes and foci, a logistic mixed model shows it is a significant predictor for phrasing ($\beta = 0.23; se = 0.09; p < 0.02$). The interactions with Focus on the first constituent predicted by AM-theory (e.g., since there should be no phrasing cues in the postfocal domain) did not come out significant ($\beta = 0.06; se = 0.12; p < 0.59$), nor was there an interaction with Intonation ($\beta = 0.05; se = 0.06; p < 0.36$). Pitch scaling was a reliable cue even (i) within the subset only involving utterances with focus on the first NP, where AM-theories predict neutralization ($\beta = 0.25; se = .10; p < 0.02$); and (ii), within the subset including only questions ($\beta = 0.14; se = 0.7-; p < 0.04$).

Recovering Three Dimensions. According to overlay models, the three separate dimensions should be fully recoverable from the signal. To further explore the data, we looked at residualized measures of intensity, pitch, and duration, using linear mixed models that accounted for variability due to other factors (Focus, Intonation, Position in the sentence, in the case of phrasing). The contribution of pitch to cueing phrasing is remarkably consistent across foci and intonations, with a troph on the second constituent for [AB]C vs. a peak on A[BC] (Fig. 2). The plot shows an effect of focus in declaratives in that scaling effects change when focus is placed on the B or C constituent.

Discussion. Three dimensions are partly unaffected by each other, for example there are pitch scaling effects in the postfocal domain. The significant contribution of cues where AM-theories predict neutralization, and the qualitative uniformity of the patterns where AM models predict interactions lend partial support to overlay models, despite known shortcomings (cf. Ladd, 2008). However, there are also interactions unexpected by pure overlay models, but they are different from those predicted by current theories.

1

(1) Declarative, Focus on Last Conjunct, Left-Branching:
I thought they said Lauren or Marvin and Sue arrived. But in fact they said that (Lauren$_A$ or Marvin$_B$) and Nolan$_C$ arrived.

(2) Declarative, Focus on Last Conjunct, Right-Branching
I thought they said Lauren or Marvin and Sue arrived. But in fact they said that Lauren$_A$ or (Marvin$_B$ and Nolan$_C$) arrived.

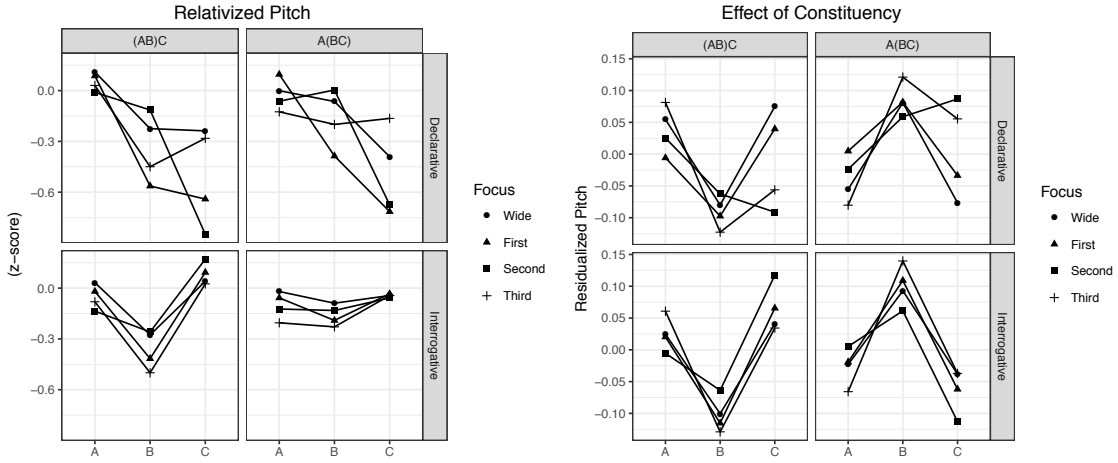


Fig. 1 (left): Z-score of mean pitch, relativized to expected values given speaker and segmental content (A,B,C) by tune (declarative vs. interrogative) and phrasing. (AB)C vs. A(BC) show heterogenuous patterns across intonations and foci.

Fig. 2 (right): Pitch residualized for factors other than phrasing (intonation, focus, position) for each target word, plotted by tune, focus, and phrasing. (AB)C and A(BC) show remarkably consistent mirror-image patterns across intonations and foci.

Beckman, M.E. 1996. The Parsing of Prosody. *Prosody and Parsing* 19:67.

Fujisaki, H. 1981. Dynamic characteristics of voice fundamental frequency in speech and singing. acoustical analysis and physiological interpretations. *KTH Department of Speech, Music and Hearing. Quarterly Status Report* .

Ladd, D. Robert. 1988. Declination and 'reset' and the hierarchical organziation of utterances. *JASA* 84.

Ladd, D. Robert. 2008. *Intonational phonology*. Cambridge: Cambridge University Press, 2nd edition.

Norcliffe, E., and T.F. Jaeger. 2005. Accent-free prosodic phrases? Accents and phrasing in the post-nuclear domain. In *Proceedings of Interspeech 2005*.

Öhmann, S. 1967. Word and sentence intonation: A quantitative model. *KTH Department of Speech, Music and Hearing. Quarterly Status Report* .

Pierrehumbert, Janet. 1980. The phonology and phonetics of English intonation. Doctoral Dissertation, MIT.

Xu, Y. 2005. Speech melody as articulatorily implemented communicative functions. *Speech Communication* 46:220–251.

Tuesday, 16:00 - 17:00
Poster Session 4, #7

Lazy speakers and distant sounds: on the role of articulatory difference in phonotactic production, acquisition, and change

Andreas Baumann[1] and Kamil Kaźmierski[2]
[1]*University of Vienna,* [2]*Adam Mickiewicz University Poznań*

Articulatory properties have been suggested to determine the success and stability of phonotactic items, i.e. sequences of sounds [1]. In this paper, we show that English diphones (i.e. sequences of two sounds) in general benefit from small articulatory differences between their elements. Empirically, we support our hypothesis by means of frequency, acquisition and diachronic data.

Articulatory difference plays a crucial role in the formation of consonant diphones. A widely established and attested principle governing phonotactics, for instance, is that of increasing/decreasing sonority of consonants (a property correlated with manner of articulation) in onset/coda position [2]. A more encompassing approach has been pursued by Dziubalska-Kołaczyk [3] who proposes a unified measure of the articulatory and perceptual difference between pairs of consonants, the so-called net-auditory distance (NAD). More specifically, she suggests that clusters with large NAD (i.e. differing substantially both in terms of manner and place of articulation) are generally preferred, since they facilitate perception.

This paper analyzes the differential effects of manner-of-articulation differences (ΔMoA) and place-of-articulation differences (ΔPoA) on the productivity, acquisition, and diachronic stability of word-final consonant diphones in English, as e.g. /nd/ in *hound* or /lz/ in *calls*. For each consonant diphone type, we calculate ΔMoA and ΔPoA as the normalized pairwise difference between ordinal sonority scores and scores that reflect the places of articulation, respectively.

In order to operationalize diachronic stability, we first extracted the semicentury-wise diphone-specific token frequencies from a compilation of historical corpora covering the past 650 years (PPCME, PPCEME, PPCMBE, COHA, COCA [4-8]). Phonological transcriptions were added manually (for PPC(E)ME) or taken from the Carnegie Mellon University Pronouncing Dictionary (for remaining corpora). For each diphone we then estimated the growth rate by fitting a logistic growth model [9] to the corresponding trajectory. The higher the growth rate, the more diachronically stable a diphone is considered. Diphone-specific age-of-acquisition (AoA) scores were extracted from a data set of AoA ratings for 30,000 English words [10]. Finally, we used the Buckeye Speech Corpus in order to determine token and type frequencies for each diphone as a measure of contemporary productivity. In total, n = 50 diphone types were taken into account.

Both ΔMoA and ΔPoA were entered into three separate generalized additive models (GAM, [11]) as predictor variables, in which logistic growth, AoA and log (token/type) frequency, respectively, figure as dependent variables. Surprisingly, the analysis shows that large articulatory differences impede diachronic growth, increase AoA and decrease productivity in speech (Figure 1). More precisely, ΔMoA (but not ΔPoA) shows a significantly decreasing impact on diachronic stability, while AoA increases with both articulatory distances. Concerning productivity, ΔPoA displays a significant impeding effect, while ΔMoA does not (an observation which interestingly goes in line with recent results from neurocognitive research, cf. [12]).

Our results thus do not corroborate, but instead indicate very much the opposite of Dziubalska-Kołaczyk's [3] claims. Small articulatory distances facilitate the production and acquisition of diphones. That is, the success and stability of diphones seems to be to a large extent determined by pressures imposed by the speaker, rather than by the listener's need to decode diphones easily.

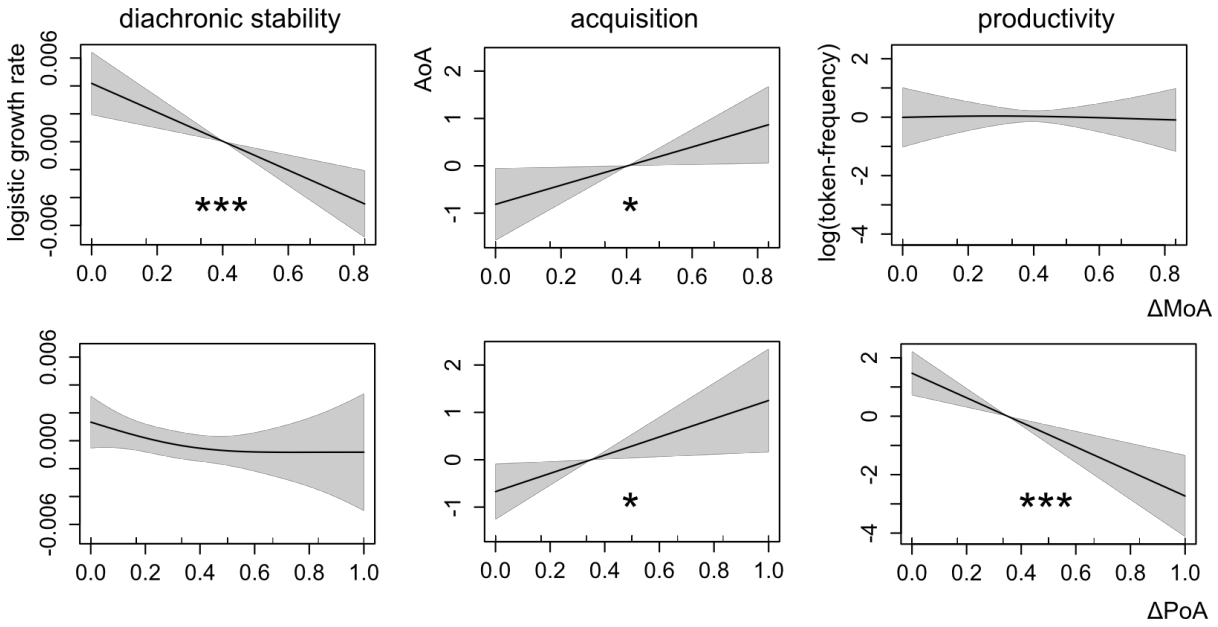


Figure 1. *The differential effects (linear predictors) of ΔMoA (upper row) and ΔPoA (lower row) on logistic growth (left), AoA (middle) and log token frequency (right). Gray areas denote 95% confidence regions. Significance code: * = p < .05, ** = p < .01, *** = p < .001.*

References

[1] Davidson, L. 2006. Phonotactics and articulatory coordination interact in phonology: Evidence from Nonnative production. *Cognitive Science* 30, 837-862.

[2] Berent, A., Steriade, D., Lennertz, T., & Vaknin, V. 2007. What we know about what we have never heard: Evidence from perceptual illusions. *Cognition* 104(3), 591–630.

[3] Dziubalska-Kołaczyk, K. 2014. Explaining phonotactics using NAD. *Language Sciences* 46, 6–17.

[4] Kroch, A., & Taylor, A. 2000. *Penn-Helsinki Parsed Corpus of Middle English*. Retrieved from http://www.ling.upenn.edu/hist-corpora/

[5] Kroch, A., Santorini, B., & Delfs, L. 2004. *Penn-Helsinki Parsed Corpus of Early Modern English*. Retrieved from http://www.ling.upenn.edu/hist-corpora/

[6] Kroch, A., Santorini, B., & Diertani, A. 2016. *The Penn Parsed Corpus of Modern British English.* University of Pennsylvania. Retrieved from http://www.ling.upenn.edu/ppche/ppche-release-2016/PPCMBE2-RELEASE-1

[7] Davies, M. 2010. *The corpus of historical American English: COHA*. Brigham: BYE, Brigham Young University.

[8] Davies, M. 2008. *The corpus of contemporary American English (COCA): 450 million words, 1990-2012*. Brigham Young University.

[9] Denison, D. 2003. Log(ist)ic and simplistic S-curves. In Raymond Hickey (ed.), *Motives for Language Change*. Cambridge: Cambridge University Press, 54–70.

[10] Kuperman, V., Stadthagen-Gonzalez, H., Brysbaert, M. 2012. Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods* 44, 978–990.

[11] Wood, S. 2006. *Generalized Additive Models: an introduction with R*. Boca Raton, Florida: Chapman & Hall/CRC.

[12] Mesgarani, N., Cheung, C., Johnson, K., & Chang, E. 2014. Phonetic feature encoding in human superior temporal gyrus. *Science* 343, 1006–1010.

Tuesday, 16:00 - 17:00
Poster Session 4, #8

Tracking Continuous Motor Responses as a Window into Real-Time Speech Perception

Timo B. Roettger & Mathias Stoeber
Department of Linguistics – Phonetics, University of Cologne

To sustain successful communication, listeners need to make rapid temporary inferences about what a speaker intends to communicate, even if these inferences are based on partial information. Listeners continuously integrate incoming acoustic information as the signal unfolds. To fully understand these processes involved in comprehending speech, it is imperative to use experimental techniques that operationalize the real-time integration of phonetic information. Thus, it is of fundamental importance to both constantly assess existing methodologies critically and explore new ways of illuminating these real-time processes.

Over the last two decades, eye-tracking has been a popular and successful window into speech processing. Participants usually wear a head-mounted eye-tracker that records where they look while solving tasks. Measurements such as saccadic landing positions, changes in saccadic movement, and fixation time can indicate underlying cognitive processes. This method has broadened our understanding of real-time speech perception in different domains ranging from the perception of segmental contrasts (e.g. [1]), over spoken word recognition (e.g. [2]) to the prosodic decoding of whole utterances (e.g. [3]).

However, several researchers have pointed out that the ballistic nature of oculomotor patterns is an important limitation of the eye-tracking paradigm (e.g. [4], [5]). Eye-movement data allow for approximately 3-4 data points per second characterised by ballistic "jumps" of the eye. Only by averaging over many trials can a pseudo-continuous trajectory be calculated that is interpretable as evidence for a continuous comprehension process. At any given point in time, participants either fixate an object or they do not.

This potential methodological drawback can be complemented by the measurement of manual response trajectories. Recently, it has been demonstrated that continuous nonlinear trajectories recorded from the streaming of x,y coordinates of computer mouse movements can serve as an informative indicator of cognitive processes ([4], [6], [7]). In contrast to eye-movement data, continuous motor response tracking yields between 30 and 100 data points per second and the hardware is comparably cheap. Mouse tracking has been successfully applied to linguistic research questions in the domain of lexical selection (e.g. [4]), syntactic parsing (e.g. [8]), and pragmatic inferences (e.g. [9]). More recent technical advances have expanded the original mouse tracking paradigm: For example, participants in [5] moved a wireless, arm-extended pointing device towards visual target locations on a screen. This design enabled tracking motor responses in their natural three-dimensional extension providing even richer measurements that unpack participants' real-time processing.

While continuous response tracking has been applied to a number of research areas in cognitive science, it has been somewhat neglected in the field of speech perception. This paper will introduce continuous response tracking as a tool for phoneticians and phonologists to unravel the continuity of speech perception. To that end, we will present experimental data on phenomena often discussed in the phonetic literature: Categorical perception, spoken word recognition, and post-lexical prosodic processing. Using both mouse tracking and more advanced arm extended pointing, we discuss evidence for the gradient nature of segmental speech, lexical competition during spoken word recognition, and the continuous integration of intonational information during sentence processing. We will discuss methodological advantages and possible application areas for speech scientists and we hope to convince the community that continuous response tracking is a methodology which holds great promise for low-cost but detailed exploration of fine-grained temporal aspects of speech perception in a multitude of domains.

[1] McMurray, B., Clayards, M.A., Tanenhaus, M.K., & Aslin, R.N. (2008). Tracking the time course of phonetic cue integration during spoken word recognition. *Psychonomic Bulletin & Review*, 15(6), 1064–1071.
[2] Allopenna, P.D., Magnuson, J.S., & Tanenhaus, M.K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, 38(4), 419-439.
[3] Dahan, D., Tanenhaus, M.K., & Chambers, C.G. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language*, *47*(2), 292-314.
[4] Spivey, M.J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences,* 102, 10393–10398.
[5] Dale, R., Roche, J., Snyder, K., & McCall, R. (2008). Exploring action dynamics as an index of paired-associate learning. *PLoS ONE* 3, e1728.
[6] Magnuson, J.S. (2005). Moving hand reveals dynamics of thought. *Proceedings of the National Academy of Sciences of the United States of America*, 102(29), 9995-9996.
[7] Song, J.H., & Nakayama, K. (2009). Hidden cognitive states revealed in choice reaching tasks. *Trends in Cognitive Science,* 13, 360–366.
[8] Farmer, T.A., Anderson, S.E., & Spivey, M.J. (2007). Gradiency and visual context in syntactic garden-paths. *Journal of Memory and Language,* 57, 570–595.
[9] Tomlinson, J.M., Bailey, T.M., & Bott, L. (2013). Possibly all of that and then some: Scalar implicatures are understood in two steps. *Journal of Memory and Language*, 69(1), 18-35.

Tuesday, 16:00 - 17:00
Poster Session 4, #9

Position vs. Prosody: Focus Realization in Urdu/Hindi

Farhat Jabeen
University of Konstanz, Germany

This research focuses on the interaction between structural and prosodic realization of focus in Urdu/Hindi. Urdu realizes new information focus by position and the focused constituents are placed at the immediately preverbal position (Gambhir 1984, Butt & King 1996). But the prosodic realization of new information focus in Urdu/Hindi is a controversial issue. Existing research offers conflicting evidence in this regard. Moore (1965) and Harnsberger (1999) have shown that Hindi does mark focus prosodically by using longer duration and pitch rise. Jabeen et al. (2016) and Butt et al. (2016) have also affirmed the use of prosodic cues to mark focus. However, other recent investigations (Patil et al. 2008, Choudhary 2016, Féry, Pandey & Kentner 2016) have shown that the prosodic marking of focus in Hindi is unreliable and unsystematic. To account for these varying claims, we set up a reproduction of Patil et al.'s findings to explore the realization of new information focus in Urdu. Using their data set and analysis techniques, we investigated the prosody of wide, subject, and object focus in SOV and OSV sentences. Following Patil et al., we elicited focus by using sentences in question-answer pairs based on Rooth's (1985) theory of Alternative Semantics. An example question-answer pair is given in (1).

1. mechanic=ne **kya** chəlaya? **(Object focus)**
   mechanic=Erg what drive.Perf.M.Sg
   'What did the mechanic drive?'

   mechanic=ne **taxi=ko** chəlaya
   mechanic=Erg taxi=Acc drive.Perf.M.Sg
   '(A/The) mechanic drove (a/the) taxi.'

Our data corroborates the findings of Patil et al. and shows that the prosodic realization of new information focus in Urdu is not systematic. We, however, interpret this lack of prosodic focus realization as an outcome of the interaction between word order and information structure. The immediately preverbal constituents (S**O**V, O**S**V) are not focused prosodically as they are already placed at the structural position reserved for new information focus. This also explains the lack of difference between the prosody of wide focus and immediately preverbal focused constituents. Thus we claim that new information focus in Urdu/Hindi is realized by position. Hence the prosodic focus marking is redundant.

This claim is further corroborated by a grammaticality judgment test. We used the same question-answer pairs as (1) to investigate the word order preferences of Urdu speakers to mark new information focus in a broad context as well as on subject and object nouns. The results indicate a significant preference to place the focused constituent at the preverbal position.

This analysis also explains the lack of prosodic focus on the sentence initial constituents (**S**OV, **O**SV). As Urdu/Hindi marks focus at the preverbal position, these constituents cannot carry new information focus. The sentence initial constituents are topicalized and can only be focused contrastively as per the information structure theory for Urdu/Hindi (Butt and King, Gambhir, Kidwai 2000).

To conclude, we offer a reinterpretation of existing research and explain the lack of prosodic marking of focus in Urdu/Hindi in the light of varied information structure. We also posit that a robust investigation of Urdu/Hindi prosody should take into account the interplay between word order and information structure.

References:

[1] Gambhir, V. 1981. *Syntactic Restrictions and Discourse Functions of Word Order in Standard Hindi.* Ph.D thesis, University of Pennsylvania, Philadelphia.

[2] Butt, M., & King, T. H. 1996. Structural topic and focus without movement. In Butt, M., & King, T.H. (Eds.), *Proceedings of the First LFG Conference,* CSLI Publications.

[3] Harnsberger, J. 1999. The role of metrical structure in Hindi intonation. In *South Asian Language Analysis Roundtable*, University of Illinois, Urbana-Champaign.

[4] Moore, R. 1965. *A study of Hindi intonation*. Ph.D. thesis, University of Michigan.

[5] Jabeen, F., Bögel, T., & Butt, M. 2016. Variable prosodic realizations of verb focus in Urdu/Hindi. *In Proceedings of Speech Prosody 2016.*

[6] Butt, M., Jabeen, F. & Bögel, T. 2016. Verb cluster internal wh-phrases in Urdu: Prosody, syntax and semantics/pragmatics. *Linguistic Analysis,* 40 (3-4), 445-487.

[7] Patil, U., Kentner, G., Gollrad, A., Kügler, F., Féry, C. & Vasishth, S. 2008. Focus, word order and intonation in Hindi. *Journal of South Asian Linguistics (JSAL),* 1(1), 55-72.

[8] Choudhury, A. & Kaiser, E. (2015). Interaction between prosody and focus types: Evidence from Bangla and Hindi. *Proceedings of Formal Approaches to South Asian Languages (FASAL V).*

[9] Féry, C., Pandey, P. & Kentner, G. 2016. The Prosody of focus and givenness in Hindi and Indian English. *Studies in Language* 40(2), 302–339.

[10] Rooth, M. 1965. *Association with focus*. Ph.D thesis, University of Massachusetts, Amherst.

[11] Kidwai, A. 2000. *XP-adjunction in Universal Grammar: Scrambling and binding in Hindi-Urdu*. Oxford: Oxford University Press.

Sentence-final lengthening in Hungarian: Effects of phonemic length and word length

Krepsz, Valéria – Gósy, Mária
Research Institute for Linguistics, Hungarian Academy of Sciences, Hungary

The phenomenon of final lengthening exists in various languages (e.g., [1, 2, 3, 4]) showing language-specific patterns (e.g., [5, 6, 7]). Research in 12 languages confirmed final lengthening depending on increasing word length [8]. Hungarian is an agglutinating language with rich morphology. The vowel inventory contains vowel pairs distinguished by length. The question arises whether phonemic differences in length and the word length influence the durations of sentence-final vowels as opposed to their realizations in initial and medial positions. In this study we analyzed durations of sentence initial, medial, and final, phonemically short and long vowels with various vowel qualities in words with diverse numbers of syllables (up to 5). We hypothesized that (i) target vowels in sentence-final position will be longer than in sentence initial and medial positions, (ii) lengthening will be attested both for phonemically short and long vowels, (iii) the number of syllables will influence target vowel durations irrespective of position: the more syllables the words contain, the shorter the vowels will be.

Four pairs of vowels were selected differing in phonemic length ([o, oː, i, iː] vs. [ɔ, aː, ɛ, eː]), the latter two differed also phonetically. 40 words (verbs and nouns) were selected and embedded into well-formed sentences in which the target vowels appeared in all the three positions (altogether 120 sentences). Syllable structure (closed syllables) and consonant environment were kept constant. Examples: *bor* 'wine', *adok* '/I/ give', *haladok* '/I/ proceed', *kirándulok* '/I/ go for an excursion', *kitakarítok* '/I/ clean /the room/'. The sentences were read by ten young native Hungarian speakers (mean age 22, five females in each group) read the sentences that were recorded in a sound attenuated room directly to computer. Altogether 1200 realizations were recorded in a sound attenuated room and annotated using the Praat software [9]. Duration measurements were carried out automatically using a specific script. To test statistical significance, Generalized Linear Mixed Models were used (dependent factors: measured durations; independent factors: 'phonemic length', 'vowel quality', 'number of syllables of the words', 'sentence position'; random effect: gender).

Results showed that vowel durations became significantly shorter as the number of syllables increased; however, the increase was not linear but logarithmic. Vowels were significantly longer in sentence final as opposed to medial positions, but no significant differences were found in the durations between initial and final positions. Although phonologically long vowels were significantly longer than phonologically short ones in all positions, sentence-final lengthening was more marked in the phonologically long than in the phonologically short vowels (Fig. 1). In addition, phonemically short and long vowels also show different durational patterns according to phonetic differences. Sentence-final lengthening is inversely proportional to the number of syllables in the words; the decrease in duration in sentence-final positions across increasing numbers of syllables in phonemically long vowels is significantly steeper than in phonemically short ones. No significant differences were found in the durations depending on vowel quality and gender.

As expected, both phonemically short and long vowels were significantly longer in sentence-final position than in initial and medial positions. Our study also highlighted between increasing number of syllables and sentence-final lengthening. The more syllables the words contained, the shorter the vowels were, confirming our hypotheses. We conclude that strong distinction of short and long vowels also in sentence-final positions suggests that speakers avoid violating the phonemic patterns of the vowel system. The possible reasons for our findings (breathing, cognitive factors etc.) will also be discussed.

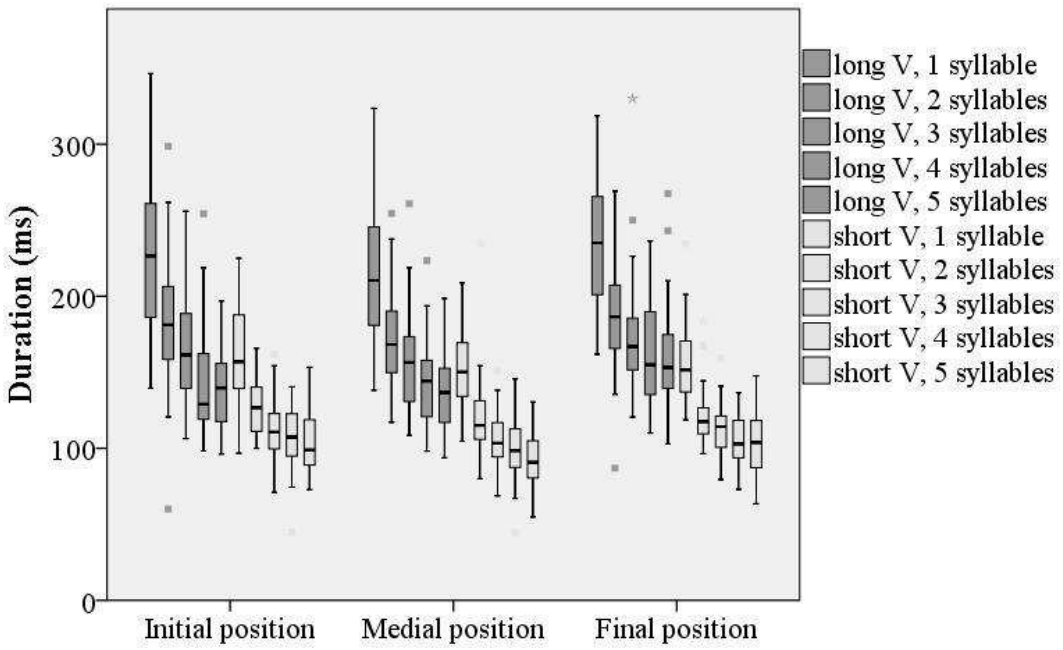


Figure 1. *Durations of phonemically short and long vowels occurring in sentence initial, medial and final positions and depending on the number of syllables of the words*.

[1] Oller, K. D. 1973. The effect of position in utterance on speech segment duration in English. *Journal of the Acoustical Society of America* 54, 1235–1247.

[2] Nakai, S., Kunnari, S., Turk, A., Suomi, K., & Ylitalo, R. 2009. Utterance-final lengthening and quantity in Northern Finnish. *Journal of Phonetics* 39, 29–45.

[3] Cho, T., Kim, J., & Kim, S. 2013. Preboundary lengthening and pre-accentual shortening across syllables in a trisyllabic word in English. *Journal of Acoustic Society of America* 133, 384–390.

[4] Den, Y. 2015. Some phonological, syntactic, and cognitive factors behind phrase-final lengthening in spontaneous Japanese: A corpus-based study. *Laboratory Phonology* 6, 337–379.

[5] Beckman, M. E. 1992. Evidence for speech rhythms across languages. In Tohkura, Y., Vatikiotis-Bateson, E. & Sagisaka, Y. (Eds.), *Speech perception, production and linguistic structure*. Oxford: IOS Press, 457–463.

[6] Berkovits, R. 1993. Utterance-final lengthening and the duration of final-stop closures. *Journal of Phonetics* 21, 479–489.

[7] Turk, A. 2007. Multiple targets of phrase-final lengthening in American English words. *Journal of Phonetics* 35, 445–472.

[8] McAuliffe, M., Sonderegger, M., & Wagner, M. 2015. A system for unified corpus analysis, applied to duration compression effects across 12 languages. Paper presented at The 15th Conference on Laboratory Phonology, Cornell University, USA.

[9] Boersma, P. & Weenink, D. 2014. *Praat: doing phonetics by computer*. Computer program.

[10] Krull, D. 1997. Prepausal lengthening in Estonian: Evidence from conversational speech. In Lehiste, I. & Ross, J. (Eds.), *Estonian prosody: Papers from a symposium*. Tallinn: Institute of Estonian Language, 136–148.

Tuesday, 16:00 - 17:00, Poster Session 4, #11

Are Consonantal Effects on F0 conditioned by Enhancement of Contrasts of Tones?

Qian Luo[1], Karthik Durvasula[1], Yen-Hwei Lin[1]

[1]*Department of Linguistics and Languages, Michigan State University*

Introduction: There are two proposed accounts for consonantal effects on vowel F0 (henceforth C-F0): (a) the *functional account* [1], [2] suggests C-F0 is an outcome of controlled articulation for enhancing a laryngeal feature; (b) the *non-functional account* may suggest the effect is due to physiological and aerodynamic properties during the articulation of the consonants [3], [4]. However, all previous literature was based on consonants, but few have considered lexical tone as a conditioning factor for the perturbation effect. This current study aims at investigating whether enhancement of tone contrasts may influence C-F0.

[5] has shown that the larger the tone inventory is in a language, the smaller the tone space at tone onset will be. For example, Cantonese with six lexical tones has a smaller onset tone space than the onset tone space of Mandarin with four lexical tones. If we take C-F0 within a certain tone category analogous to stretching one's arms in the F0 dimension: it is more likely to hit some others (=trespassing into another tone category) in a smaller space than in a larger one. Assuming that a larger tone inventory implies a smaller tone space, this indicates that a strong C-F0 effect is more likely to threaten tone contrasts in languages with larger tone inventories (e.g. Cantonese) than those with smaller ones (e.g. Mandarin).

Prediction 1: (a) The *functional account* predicts that C-F0 should be weaker in languages with larger tone inventories (e.g. Cantonese) than those with smaller ones (e.g. Mandarin), due to the more severe threat to tone contrasts in a small tone space than in a big one; (b) the *non-functional account* predicts that the difference of size of the tone inventory should not influence C-F0.

Prediction 2: Assuming the tone contrast only needs to be enhanced within L1 or L2, but not across L1 and L2, (a) the *functional account* predicts that the magnitudes of C-F0 are different between L1 and L2 for bilingual speakers, given L1 and L2 have different sizes of tone inventories; (b) the *non-functional account* predicts that the magnitudes of C-F0 should not be influenced by different sizes of tone inventories.

Experimental Design: Mandarin speakers (4M, 2F) and Cantonese speakers (2M, 4F) participated in the production experiments. All Cantonese participants also spoke native-like Mandarin and continued to participate in the Mandarin experiment. All stimuli followed a CV template. The four tonal contexts were Cantonese High-Level and Mid-Rising tones, and Mandarin High-Level and Mid-Rising tones. Initial consonants include aspirated stops [t^h, k^h, p^h], unaspirated stops [t, k, p] and sonorants [n, m, l]. F0 values following sonorants were the baseline for C-F0 [6], [7]. The mean F0 values within the first 20 ms of the vowel were analyzed.

Results: As shown in Figure 1, F0 is significantly higher after the unaspirated and aspirated obstruents than after the sonorant baseline in the High-Level tone contexts in both Cantonese and native Mandarin. F0 is significantly lower after aspirated obstruents than after unaspirated obstruents and sonorants in the Mandarin Mid-Rising tone context. No other finding is statistically significant. Comparing the F0 difference between the lowest/highest mean F0 after obstruents and the sonorant baseline, Cantonese has a higher F0 difference than Cantonese Mandarin, and higher than native Mandarin.

Conclusion and Discussion: The results do not support predictions 1 and 2 by the functional account, for Cantonese does not have a weaker C-F0 than Cantonese Mandarin and native Mandarin. We may infer that the difference in C-F0 in the three language contexts is more likely to be due to different physical properties in articulation. Further studies are needed to directly investigate what physical properties are responsible for magnitudes and directions of C-F0.

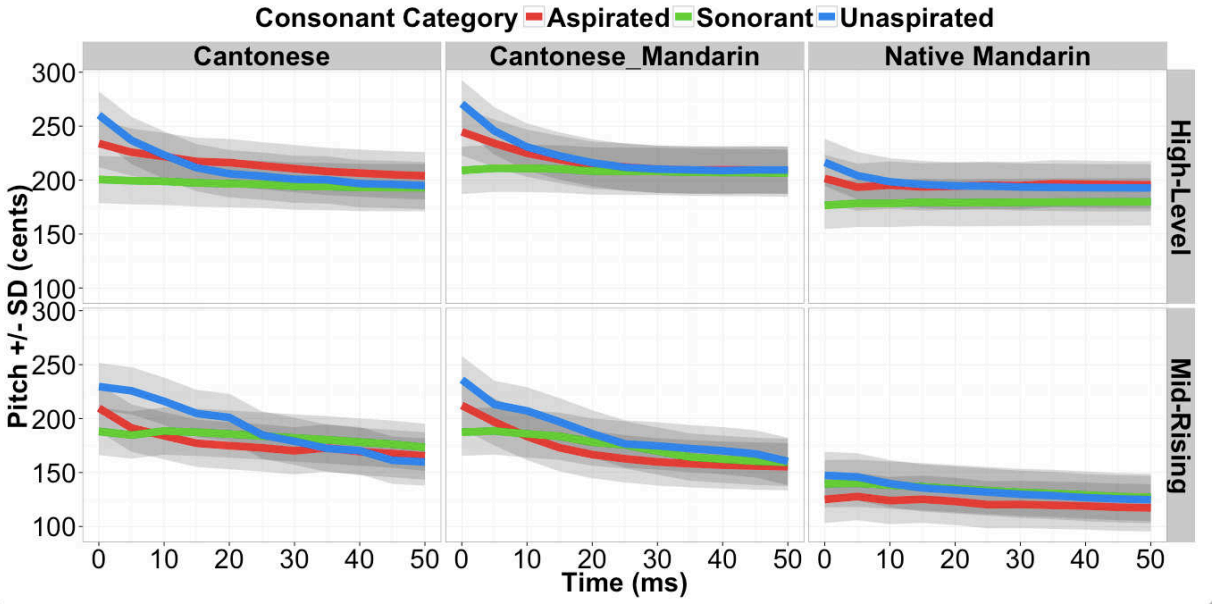


Figure 1

References

[1] J. Kingston and R. L. Diehl, "Phonetic Knowledge," *Language*, vol. 70, no. 3, pp. 419–454, 1994.

[2] J. Kingston, "Segmental influences on F0: Automatic or controlled?," in *Tones and Tunes*, vol. 2, C. Gussenhoven and T. Riad, Eds. Berlin: Mouton de Gruyter, 2007, pp. 171–210.

[3] D. H. Whalen and A. G. Levitt, "The universality of intrinsic F0 of vowels," *Journal of Phonetics*, vol. 23, no. 3, pp. 349–366, 1995.

[4] B. Connell, "Tone languages and the universality of intrinsic F 0: evidence from Africa," *Journal of Phonetics*, vol. 30, no. 1, pp. 101–129, 2002.

[5] J. A. Alexander, "The theory of adaptive dispersion and acoustic-phonetic properties of cross-language lexical-tone systems," PhD Dissertation, Northwestern University, 2010.

[6] H. M. Hanson, "Effects of obstruent consonants on fundamental frequency at vowel onset in English," *The Journal of the Acoustical Society of America*, vol. 125, no. 1, p. 425, 2009.

[7] J. Kirby and D. R. Ladd, "Stop voicing and f0 perturbations: Evidence from French and Italian," in *Proceedings of the 18th International Congress of Phonetic Sciences, Glasgow, Scotland*, 2015.

Tuesday, 16:00 - 17:00
Poster Session 4, #12

Normalizing versus initializing cues to pitch (mis-)matching in conversation

Margaret Zellers[1] & Antje Schweitzer[2]
[1]*Department of Linguistics: English, University of Stuttgart, Germany*
[2]*Institute for Natural Language Processing, University of Stuttgart, Germany*

Pitch has been reported by many researchers to serve as a contextualization cue in conversation; that is, it provides information which allows listeners to infer meaning of utterances beyond their lexical content and make "situated interpretations" (Gumperz, 1992). The "situation" in which these interpretations occur is the surrounding turns; the position and the form of a conversational turn are both relevant in understanding what a speaker is doing (Schegloff, 1993, 2002). Among many other functions, pitch can help communicate discourse-internal factors such as solidarity between speakers (Couper-Kuhlen & Selting, 1996) and (dis)agreement (Ogden 2006). Furthermore, speakers have been shown to accommodate or converge prosodically to one another during some phases of discourse, although the exact mechanisms underlying such accommodation are still unclear (Gallois & Callan, 1988; Giles et al., 1991; Pickering & Garrod, 2004; Lewandowski, 2012; de Looze et al., 2014).

Listeners are able to quickly perceive and adjust (if necessary) to their interlocutor's speech, and in particular prosodic patterns. Perceptual studies have indicated that listeners are able to reliably identify the location of fundamental frequency (F0) values within an individual speaker's range (Honorof & Whalen, 2005). While this may be due in part to listeners' expectations regarding average F0 for different speaker sexes (Bishop & Keating, 2012), voice quality also contributes to the perception of pitch height, and manipulating voice quality can manipulate listeners' judgments about the location of F0 values within a speaker's pitch range (Kuang & Liberman, 2016). Thus it appears that even without conversational context, listeners have some ability to rank pitch as high or low; that is, they normalize for a speaker (Ladd, 2008). However, researchers studying pitch specifically in the context of conversation have argued strongly for an approach in which a turn's pitch must be judged relative to what has come before instead of (only) relative to the speaker's range; Ladd (2008) calls this an initializing approach. Recent research into the production of prosody in adjacent turns has provided support for both normalizing and initializing views. Gorisch et al. (2012) find that backchannels have pitch contours that closely match the pitch contour at the end of the turn they follow, which is difficult to account for using normalization. Sicoli et al. (2014), on the other hand, report that the function of turns with question structure can be identified by whether they have high or low initial pitch (i.e. not near the speaker's median pitch, and thus considered to mismatch with the context), without reference to the preceding turn's pitch. Despite conflicting reports based on production data, however, to our knowledge no study has been carried out investigating how listeners rank prosodic matching or mismatching.

We will report the results of a study in which listeners rank pitch of adjacent conversational turns from the GECO Corpus (Schweitzer & Lewandowski, 2013) using both initializing and normalizing paradigms. In one portion of the experiment, listeners hear two adjacent turns and rank whether the second turn has relatively the same, higher, or lower pitch than the first. In the other portion of the experiment, listeners hear single turns and rank them as having high, normal, or low pitch. We investigate the relationship of listeners' rankings in the two portions, for example, whether a ranking as high pitch for both speakers relative to their individual ranges will also equate to pitch "matching" across adjacent turns. We also consider the relationship of these rankings to the acoustic characteristics of the turns. The results of this experiment will shed light on the relationship between pitch normalization and perceived pitch matching in adjacent conversational turns. This in turn will help give a better understanding of how social actions such as agreement and disagreement are accomplished in conversation, as well as to what extent we should expect prosodic accommodation to involve precise matching versus general trends.

[1] Gumperz, J.J. (1992) Contextualization and understanding. In Duranti, A. & C. Goodwin (eds.) *Rethinking context: language as an interactive phenomenon*. Cambridge: Cambridge University Press, pp. 229-252.
[2] Schegloff, E.A. (1993) Reflections on quantification in the study of conversation. *Research on Language and Social Interaction* 26: 99-128.
[3] Schegloff, E.A. (2002) Accounts of conduct in interaction: Interruption, overlap and turn-taking. In J. H. Turner (ed). *Handbook of Sociological Theory*. New York: Plenum, pp. 287–321.
[4] Couper-Kuhlen, E. & Selting, M. (1996) Towards an interactional perspective on prosody and a prosodic perspective on interaction. In Couper-Kuhlen, E. and Selting, M. (eds.) *Prosody in Conversation: Interactional Studies*. Cambridge: Cambridge University Press, pp. 11–56.
[5] Ogden, R. (2006) Phonetics and social action in agreements and disagreements. *Journal of Pragmatics* 38(10): 1752-1775.
[6] Gallois, C., & Callan, V. J. (1988) Communication accommodation and the prototypical speaker: Predicting evaluations of status and solidarity. *Language & Communication* 8(3): 271-283.
[7] Giles, H., Coupland, J., & Coupland, N. (1991) *Contexts of accommodation*. Cambridge, UK: Cambridge University Press.
[8] Pickering, M.J., & Garrod, S. (2004) The interactive-alignment model: Developments and refinements. *Behavioral and Brain Sciences* 27(2): 212-225.
[9] Lewandowski, N. (2012) *Talent in non-native convergence*. PhD dissertation, University of Stuttgart, Germany.
[10] De Looze, C., Scherer, S., Vaughan, B., & Campbell, N. (2014) Investigating automatic measurements of prosodic accommodation and its dynamics in social interaction. *Speech Communication* 58: 11-34.
[11] Honorof, D.N. & Whalen, D.H. (2005) Perception of pitch location within a speaker's F0 range. *Journal of the Acoustical Society of America* 117(4): 655.
[12] Bishop, J. & Keating, P. (2012) Perception of pitch location within a speaker's range: fundamental frequency, voice quality and speaker sex. *Journal of the Acoustical Society of America* 132(2): 1100-1112.
[13] Kuang, J. & Liberman, M. (2016) Pitch-range perception: the dynamic interaction between voice quality and fundamental frequency. *Proceedings of 17th Interspeech, San Francisco, CA, USA,* 1350-1354.
[14] Ladd, D.R. (2008) *Intonational phonology*. Cambridge, UK: Cambridge University Press.
[15] Gorisch, J., Wells, B., & Brown, G.J. (2012) Pitch contour matching and interactional alignment across turns: An acoustic investigation. *Language and Speech* 55(1) : 57-76.
[16] Sicoli, M.A., Stivers, T., Enfield, N.J., & Levinson, S.C. (2014) Marked initial pitch in questions signals marked communicative function. *Language and Speech* 58(2): 204-223.
[17] Schweitzer, A. & Lewandowski, N. (2013) Convergence of articulation rate in spontaneous speech. *Proceedings of 14th Interspeech, Lyon, France*, 525-529.

The intonation of alternative questions in French

Elisabeth Delais-Roussarie[1] and Giusy Turco[2]
[1]*UMR 7110 – LLF (Laboratoire de Linguistique Formelle) &Université Paris 7*
[2] *UMR 7018-LPP (Laboratoire de Phonétique & Phonologie) & Université Paris 3*

The intonational pattern that characterizes alternative questions in French (see example (1) below) can be adequately accounted for by taking into account the modality and information structure associated with the utterance. It seems that syntax-phonology mapping rules cannot fully explain the intonational differences observed in alternative questions and statements with disjunctions. To our mind, it is thus necessary to argue for a specific class of boundary tone, the *illocutionary boundary tone*, whose form indicates sentence modality, and its position ground/ focus articulation. The aim of this paper is thus twofold: (i) from an analysis of the intonation pattern observed in alternative questions, the necessity of such an illocutionary boundary tone is explained, and (ii) from investigating the use of such tone type, its phonological and phonetic characteristics are described.

A pilot production study was run in order to compare the prosody of alternative questions in (1) with the one occurring in assertions with a disjunction in (2a).

(1) a. Il veut aller à Madrid, à Amsterdam ou à Berlin?
(2) a. Il veut aller à Madrid, à Amsterdam ou à Berlin
b. Il va accueillir Nicolas ou il va aller en vacances chez Julien.

The acoustic analysis of the data showed that the rising contour occurring at the end of the first and second conjunct (i.e. *à Madrid* or *à Amsterdam* in (1) and (2a)) was different between the two types of sentences (contra [3]), regardless of number and size of the non-final conjuncts: larger pitch range and steeper rise in questions than assertions (see fig. 1 and fig 2 respectively).

The syntactic and semantic analyses of alternative questions, according to which each conjunct corresponds to a clause (see [2] and [5] among others), could provide an analysis in terms of syntax-phonology mapping to account for the realizational differences observed: an IP boundary is located after each conjunct in questions, whereas it is a mere *ip* boundary in assertions, as in (3).

(3) a. Question: [il veut aller à Madrid]$_{IP}$ [à Amsterdam] $_{IP}$ [ou à Berlin] $_{IP}$
b. Assertion [{il veut aller à Madrid}$_{ip}$ {à Amsterdam}$_{ip}$ {ou à Berlin}$_{ip}$]

The analysis in (3), however, does not seem to hold for at least two reasons:

- since the realization of the IP boundary at the end of non-final conjuncts in (3a) derives primarily from the syntactic status of the conjunct (a clause vs. a syntactic phrase in assertions), we should also observe an IP boundary in an assertion where clauses are coordinated as in (2b). However, this is not what was observed in our data;
- despite the steeper and larger rise, syllabic duration was shorter and pauses were less marked at the end of non-final conjuncts in questions than in assertions (see, for instance, the pause marked with '#' symbol in fig. 3). The phrasing obtained in (3) cannot account for such discrepancy.

In order to resolve the contradiction, we assume that the boundary tone occurring at the end of the first conjunct is an *illocutionary boundary tone* whose function is to indicate sentence modality and ground focus articulation (see [4] and [1] for similar tonal events). Such a tone displays several characteristics: (i) it occurs once per utterance, but not necessarily in final position; and (ii) it can be spread on any metrically strong position occurring on its right.

Based on the results of our production experiment and on previous work on focus-marking in French (see [1]), we will argue for a parallel in terms of phrasing and text-tune association between alternative questions and narrow focus in questions and assertions.

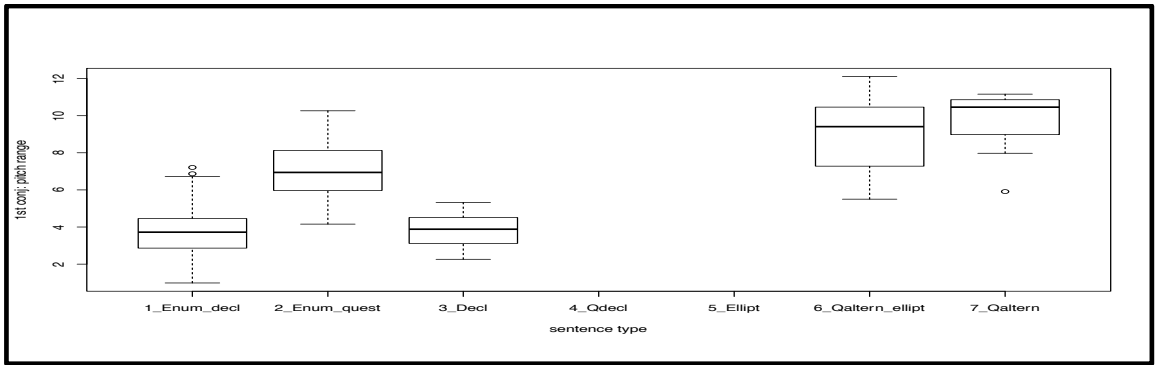


Fig. 1: Pitch range at the end of the first conjunt in alternative questions (2_enum_quest, 6_Qaltern_ellipt and 7_Qaltern) and assertions (2_enum_quest and 3_decl)

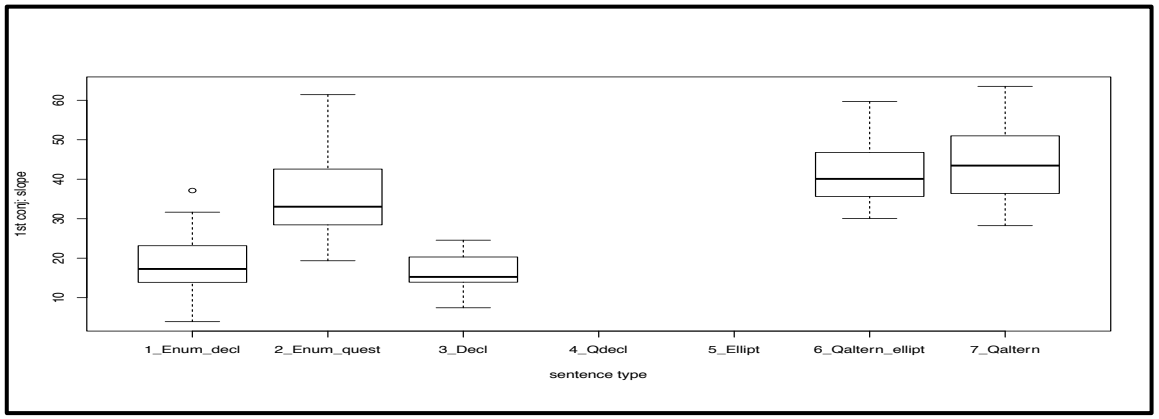


Fig. 2: Slope of the rise occurring at the end of the first conjunct in alternative questions (2_enum_quest, 6_Qaltern_ellipt and 7_Qaltern) and assertions (2_enum_quest and 3_decl)

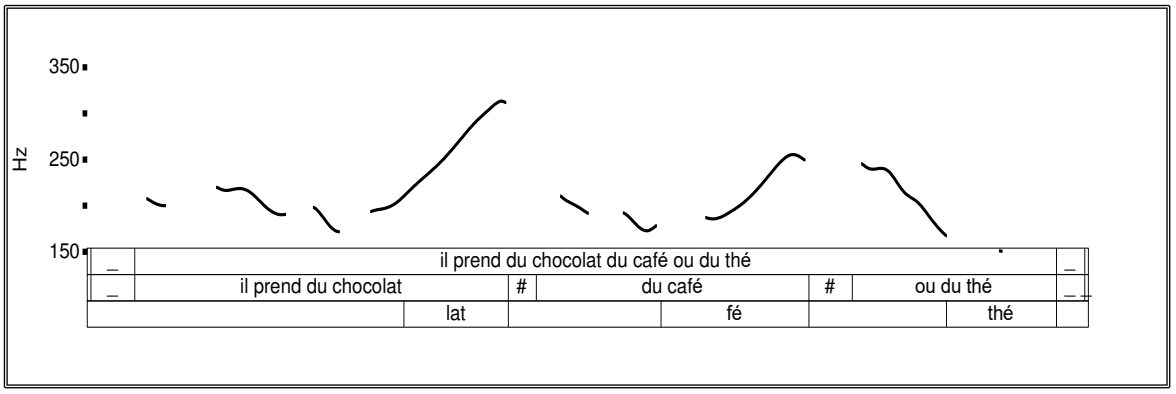


Fig. 3: Pitch track example of the French assertion *il prend du chocolat, du café ou du thé* (He takes chocolate, coffee or tea) with a pause after *chocolat* and *café.*

Selected References

[1] Beyssade C., Delais-Roussarie E., Doetjes J.S., Marandin J.-M. & Rialland A. (2004), Prosody and information in French. In: Corblin F., Swart H. de (Eds.) *Handbook of French Semantics*. Stanford: CSLI publications. 477-500

[2] de Cornulier Benoît (1982) Sur le sens des questions totales et alternatives. *Langages*, 67, 55-109.

[3] Di Cristo, A. (2016). *Les musiques du français parlé*. Berlin, Boston: De Gruyter Mouton.

[4] Martin, P. (1978): Questions de phonosyntaxe et de phonosémantique en français, *Linguisticae Investigationes* II - 1978 : 93-126.

[5] Truckenbrodt, H. (2013) An analysis of prosodic F-effects in interrogatives: prosody, syntax and semantics. *Lingua*, 124, 131-175.

Tuesday, 16:00 - 17:00
Poster Session 4, #14

Perceptual Category Adaptation: An Index of Cross-Language Coupling

Stephen Tobin
Universität Potsdam, Haskins Laboratories

Perceptual adaptation has received much attention in purely perceptual investigations [1, 2]. However, influential theories of speech have posited a direct relation between production and perception [3, 4]. In some recent investigations, consideration has been given to both [5, 6]. Among their results is the observation of perceptual adaptation as a corollary to adaptations in speech production [7]. In the present investigation I extend this line of perceptual research explicitly informed by production, considering L1 effects on L2 perceptual adaptation.

I report findings of perceptual adaptation in voiceless stops following a word-shadowing task. Spanish-English bilinguals, Korean-English bilinguals, and monolingual English controls participated in an experiment designed to induce and measure change in phonetic categories. Spanish voiceless stop are shorter (~15 ms) than those of English (~70 ms) and Korean aspirated stops are longer (~100 ms) [8], such that these L1 influences can be detected in L2 speech. All participants completed a perceptual task in which they responded to tokens from a 40-step *girl* to *curl* VOT continuum. They gave goodness ratings on a scale from 1 to 7, judging how good an exemplar of the word *curl* each token was [9]. This task typically yields curves that increases rapidly at the voicing boundary and decreases more gradually at longer, hyper-aspirated VOT values, the most highly rated tokens (best-exemplar range) lying between these extremes (Fig. 1). The task was presented in six blocks, the first of which was a baseline. The remaining five (test) blocks were interleaved with a shadowing task, intended to induce adaptation, in which participants repeated 40 [k]-initial English monosyllables spoken by a female native speaker of English.

Statistical analyses were conducted using General Additive Mixed Modelling (GAMM; [10-12]) in *R* [13]. This technique is similar to mixed polynomial regression (growth curve analysis, [14]), differing in that smoothing spline functions are used to fit continuous variables, instead of polynomials, in which higher order polynomials can yield artefactual fluctuations at window edges. The analysis yielded significant interactions of Language, Task and VOT step, such that the Korean group's ratings significantly increased in the test task at all but the shortest VOT steps and especially at longer VOTs (>125 ms; Fig. 2). The Spanish group's goodness curves became flatter in the test task, exhibiting a significant reduction in ratings in the best-exemplar range and a significant increase in VOT ratings at the voiced end of the continuum. The English group's goodness curves showed effects akin to selective adaptation: tokens near the voicing boundary were rated lower in the test task while longer VOT steps were rated higher. I propose that the effects observed in the bilingual groups are manifestations of the cross-language coupling between L1 and L2 categories, increased ratings of the Spanish group at shorter VOTs demonstrating affinity between the short positive-lag voiceless stops of Spanish and the English stops, and increased ratings of the Korean group at longer VOTs demonstrating affinity between the long positive-lag stops of Korean and those of English. Bilingual participants are surely familiar with accented voiceless stops with VOTs in these ranges, either in their own speech or in that of other native speakers of their L1.

The analysis also yielded significant interactions of Language and Task with Time for the Korean group in the baseline and test tasks, and for the Spanish group in the test task. In all of these cases, the peak ratings appear to decline gradually during the first two thirds to three quarters of the task before gently rising again at the very end of the task. This effect is not present in the Spanish group's baseline task, although a gradual decline is still visible (Fig. 3). No such effect was visible in the either of the English control group's tasks. This likely reflects greater stability of phonetic categories in the English group (vs. greater malleability of categories among the bilingual groups) as well as greater homogeneity of English proficiency.

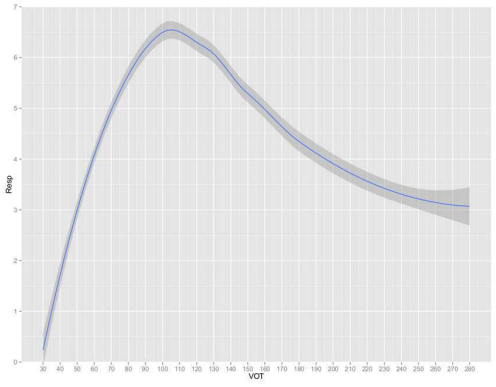


Figure 1. *Sample Goodness Judgment Curve*.

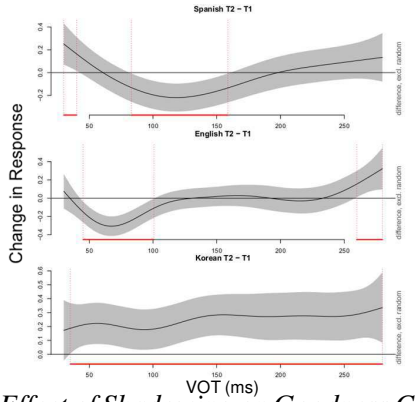


Figure 2. *Effect of Shadowing on Goodness Curves*.

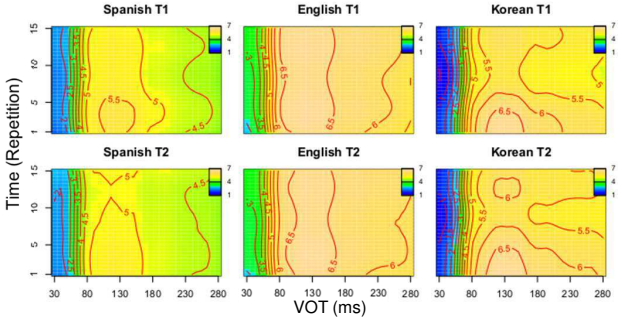


Figure 3. *Goodness Contour Plots over Time.*

[1]Kuhl, P., Conboy B., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., Nelson, T. (2008). Phonetic learning as a pathway to lang.: new data and native lang. magnet theory expanded (NLM-e). *Ph. Tr. R. Soc. B. 363*, 978-1000.

[2] Werker, J. F., & Tees, R. C. 1984. Phonemic and phonetic factors in adult cross-language speech perception. *JASA, 75*, 1866–1878.

[3] Goldstein, L. & Fowler, C. 2003. Articulatory phonology: a phonology for public language use. In N. Schiller and A. Meyer (eds.) *Phonetics and Phonology in Language Comprehension and Production.* Berlin: Mouton de Gruyter. pp. 159–208.

[4] Liberman, A. & Whalen, D., 2000. On the relation of speech to language. *TiCS, 4*(5), 187-96.

[5] Fowler, C. Brown, J., Sabadini, L. & Weihing, J. 2003. Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks. *JML. 49*(3), 396-413.

[6] Roon, K. & Gafos, A. 2015. Perceptuo-motor effects of response-distractor compatibility in speech: beyond phonemic identity. *Psychonomic Bulletin and Review, 22*(1), 242-250.

[7] Shiller, D., Sato, M., Gracco, V. & Baum, S. 2008. Perceptual recalibration of speech sounds following speech motor learning. *JASA, 125*(2), 1103-1113.

[8]Lisker, L. & Abramson, A. 1964. A cross-language study of voicing in initial stops: Acoustic measurements. *Word, 20*(3).

[9] Miller, J. 1994. On the internal structure of phonetic categories: a progress report. *Cognition, 50*(1), 271-285.

[10] Winter B., & Wieling, M. 2016. How to analyze linguistic change using mixed models, Growth Curve Analysis and Generalized Additive Modeling. *Journal of Language Evolution, 1*(1), 7-18.

[11] Wood, S. (2011) Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B) 73*(1):3-36.

[12] van Rij J., Wieling M., Baayen, R. and van Rijn, H. (2016). "itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs." R package version 2.2.

[13] R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

[14] Singer J., & Willett, J. 2003. *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence.* OUP.

Speaker variability and accomodation processes: the case of Bari Italian question intonation

Michelina Savino
Dept. of Education, Psychology, Communication, University of Bari - Italy

A growing body of research is devoted to investigating prosodic variability in relation to a number of factors, including accomodation. In this respect, question intonation deserves particular attention, as it has been observed that interrogatives are intonationally more prone to variability than statements [1], and that prosodic adaptation in language contact involves much strongly interrogatives than declaratives [2]. This paper deals with the intonational variability of polar questions in Bari Italian as connected to prosodic accomodation processes. In a preliminary study, we analysed the F0 contour of unbiased yes-no questions produced by five female Bari Italian (henceforth BI) and five female Lecce Italian (henceforth LI) speakers interacting in pairs. Note that in these two varieties, unbiased yes-no questions are prototypically realised with two completely different tunes [3]: BI has a rising nuclear accent followed by a low boundary (L+H*L-L%), even though a rising L-H% variant is frequently encountered; LI has a terminal rise preceded by an accentual low or falling target, described with either L*L-H% [3] or H+L*L-H% [4]. Results of this first study revealed a clear tendency by LI speakers to use their prototypical contour much more systematically than their BI game partners [5]. All BI speakers showed a higher production of L+H* followed by a rising instead of a falling boundary. Moreover, in some of the BI speakers variability included the use of (H+)L*L-H% contours, i.e. those typical of their LI counterparts. These results were interpreted as manifestation of prosodic adaptation in asymmetric terms, i.e. BI speakers accomodating their interrogative intonation to their LI partners' by either a) "selective imitation" of the rising L-H%, and/or b) copying of the (falling-)rising LI question contour. In order to verify that variability in question intonation showed by BI speakers was indeed triggered by an underlying accomodation process, and not possibly reflecting speaker-specific behaviour, in this study we asked the same BI subjects to participate to another session, this time interacting with a BI partner. As in the BI-LI sessions, subjects were involved in the popular game called "Guess who?", where each participant tries to guess which character is drawn on each of her partner's cards by asking information exclusively via yes-no questions. Around 500 recorded questions produced during the BI-BI sessions were manually segmented and intonationally annotated using Praat [6]. F0 contours produced in both conditions (BI-BI and BI-LI sessions) were also semi-automatically smoothed and interpolated by using an adapted version of Mausmooth [7]. As shown in Table 1, comparison of F0 contour types produced by BI speakers in both BI-BI and BI-LI sessions reveal that the production of the rising L-H% after the L+H* nuclear accent is prevailing across BI subjects in both BI-LI and BI-BI conditions. This seems to indicate that a massive use of the L-H% variant in BI speakers during the BI-LI sessions cannot necessarily be interpreted as a sort of "selective imitation" of the LI prototypical contour. On the other hand, we found F0 excursion of the L-H% in L+H*L-H% tunes as a possible relevant feature involved in the prosodic accomodation process, since two out of the five BI speakers exibited a significantly higher F0 excursion size of the L-H% when interacting with the LI than with the BI partner (see Table 2). Comparative results in Table 1 also show that the three BI speakers producing a considerable amount of (H+)L*L-H% contours during the BI-LI sessions (speakers 2, 3, 4), exibited some use of this tune also when interacting with a BI partner (see example in Figure 1). This seems to preliminarly indicate that Italian speakers, given the regular contact with speakers of geographically neighbouring yet intonationally different varieties, can incorporate a larger variability within their intonational inventory (especially for questions), which could be suitably "exploited" in accomodation processes.

		L+H*L-L%	L+H*L-H%	(H+)L*L-H%	total # of questions
BI speaker_1 interacting with	a LI speaker	25%	75%	-	93
	a BI speaker	9%	91%	-	97
BI speaker_2 interacting with	a LI speaker	21%	22%	57%	98
	a BI speaker	49%	39%	12%	87
BI speaker_3 interacting with	a LI speaker	13%	60%	27%	99
	a BI speaker	24%	73%	3%	109
BI speaker_4 interacting with	a LI speaker	3%	37%	60%	109
	a BI speaker	16%	46%	38%	102
BI speaker_5 interacting with	a LI speaker	15%	78%	7%	71
	a BI speaker	19%	81%	-	79

Table 1. *Distribution (in %) of the intonation contour types produced by each of the five BI speakers in each of the two game sessions, namely when interacting with a LI vs a BI partner.*

		F0 excursion (mean values in st) of L-H% in L+H*L-H%	t values (two-tailed paired t-test)
BI speaker_2 interacting with	a LI speaker	2.5 (1.3)	-2.77
	a BI speaker	1.7 (0.7)	$p < .01$
BI speaker_3 interacting with	a LI speaker	2.3 (0.8)	-3.47
	a BI speaker	1.9 (0.8)	$p < .001$

Table 2. *Mean values (in semitones, sdev in brackets) of F0 excursion size of the L-H% terminal rises in L+H*L-H% contours produced by BI speakers in both BI-LI and BI-BI game sessions. Results reported only when statistically significant (i.e. for BI speakers 2 and 3).*

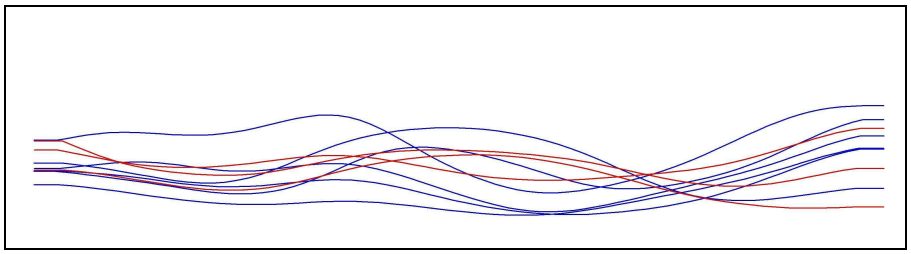


Figure 1. *Utterance-long time normalised F0 contours (vertical range: 75-400) of all realisations of the interrogative utterance 'è UOmo?' (is it MAN?) produced by one of the BI speakers when interacting with a LI partner (blue line) and with a BI partner (red line) in two different sessions.*

[1] Cangemi, F. & Grice, M. 2016. The importance of a distributional approach to categoriality in autosegmental-metrical accounts of intonation. *Journal of the Association of Laboratory Phonology* 7(1): 9, 1–20.

[2] Romera M. & Elordieta, G. 2013. Prosodic accomodation in language contact : Spanish intonation in Majorca. *International Journal of the Sociology of Language*, 221, 127–151.

[3] Savino, M. 2012. The intonation of polar questions in Italian : Where is the rise ? *Journal of the International Phonetic Association,* 42(1), 23–48.

[4] Gili Fivela, B. Avesani, C., Barone, M., Bocci, G., Crocco, C., D'Imperio, M., Giordano, R., Marotta, G., Savino, M. & Sorianello, P. 2015. Intonational phonology of the regional varieties of Italian. In Frota, S. & Prieto, P. (Eds.), *Intonation in Romance*. Oxford: Oxford University Press, 140–197.

[5] Savino, M. 2014. Prosodic adaptation in game interaction between speakers of two Italian varieties. *Proceedings of the 10th International Seminar on Speech Production*, 375–378.

[6] Boersma, P. & Weenink, D. 2008. *Praat: doing phonetics by computer*. Computer program.

[7] Cangemi, F. 2015. *Mausmooth*. Retrievable online at http://phonetik.phil-fak.uni-koeln.de/fcangemi.html

Can EPG contacts explain intraoral pressure shapes in voiced and voiceless stops in Turkish? Evidence from Generalized Additive Mixed Models

Özlem Unal-Logacev[1], Susanne Fuchs[2], Leonardo Lancia[3]
[1]*Istanbul Medipol University, Istanbul,* [2]*Leibniz Centre General Linguistics, Berlin,*
[3] *Laboratoire de Phonétique et Phonologie (CNRS, Sorbonne Nouvelle), Paris*

Motor equivalence is a basic principle in motor control describing the capacity of the motor system to achieve the same goal with different underlying mechanisms [8, 9]. We follow Fuchs [3] who proposed to study the voicing contrast within this perspective. Maintaining or diminishing voicing during oral closure is a crucial auditory cue [11] distinguishing between phonologically voiced and voiceless obstruents in languages with a true voicing contrast. Maintaining voicing during oral closure as in phonologically voiced stops requires a transglottal pressure drop between subglottal and intraoral pressure. To guarantee such a pressure drop, the oral cavity needs to be enlarged to keep intraoral pressure low [6, 7, 14]. Several cavity enlargement manoeuvres have been reported in the literature. Among them are laryngeal lowering [2], tongue retraction [13], jaw lowering [5] and others [1, 12, 10]. If no cavity enlargement manoeuvres are realized or laryngeal-oral timing is changed, intraoral pressure rises up quickly during closure (steep intraoral pressure slope) and voicing dies out. In the current study we test how much tongue palatal contact patterns predict the intraoral pressure rise in phonologically voiced versus voiceless stops, i.e. to what extent tongue placement accounts for cavity enlargement. The originality of our approach lies in the combination of articulatory and aerodynamic measures investigating their mutual dependence through Generalized Additive Mixed Models (GAMMs). Moreover, we have carried out such an analysis for an under-investigated language, Turkish. Simultaneous recordings were carried by means of electropalatography, a piezoresistive intraoral pressure sensor [4] and acoustics for six speakers of Standard Turkish. All target words were bisyllabic, preceded by a word ending with a vowel and contained one of the four target sounds /t, d, tʃ, dʒ/ in word-initial position. The voiced affricate is particularly interesting in this respect, because it is rather rare in the sounds of the world's languages [14] and the articulatory-aerodynamic relations have not been studied empirically. For EPG data the percent of overall contact (PC) was calculated. We considered for intraoral pressure and PC data within the interval from the end of the preceding vowel to the intraoral pressure peak. Intraoral pressure was defined as the dependent variable while voicing (voiced vs. voiceless) and manner (stops and affricates) and their interaction served as fixed effects. The model included manner and voice specific smooth predictors for PC. The random structure included smooth terms for percent of contact by manner and by voice as well as speaker-specific random smooths for voice, manner and target words. Results reveal that an increase in the percent of palatal contact coincides with an increase in intraoral pressure. This effect is more positive in voiceless stops and affricates in comparison to voiced ones and supports the idea that intraoral pressure rises as a direct consequence of oral closure in voiceless stops. In voiced stops and affricates, tongue palatal contacts partially account for the intraoral pressure rise, but cannot explain it alone. Furthermore, the estimated differences for an effect of palatal contact patterns on pressure show significant differences between voiced and voiceless stops/affricates shortly after the beginning of closure. However, the estimated differences for PC on intraoral pressure did not differ between voiced stops and affricates, but between voiceless stops and affricates. We suggest that GAMMs are a promising technique to investigate continuous articulatory and aerodynamic time series to understand their interaction in more detail. Based on our analysis we conclude that in Turkish not only the tongue palatal contacts are responsible for the maintenance of voicing during closure, but some additional articulatory manoeuvre should also be involved.

References:

[1] Bell-Berti, F. & Hirose, H. 1975. Palatal activity in voicing distinctions: A simultaneous fiberoptic and electromyographic study. *J. Phon.*, 3, 1, 69-74.

[2] Ewan, W. G. & Krones, R. 1974. Measuring larynx movement using the thyroumbrometer. *J. Phon.* 2, 327-335.

[3] Fuchs, S. 2005. *Articulatory correlates of the voicing contrast in alveolar obstruent production in German* (ZAS Papers in Linguistics). PhD thesis at QMUC Edinburgh.

[4] Fuchs, S. & Koenig, L.L. 2009. The articulation and aerodynamics of voiceless consonants, clusters, and affricates of German. *J. Acoust. Soc. Am.*, 126 (4), 1988-2001.

[5] Fujimura, O. & Miller J.E. 1979. Mandible height and syllable-final tenseness. *Phonetica* 36, 263-272.

[6] Koenig, L. L., & Lucero, J. C. 2008. Stop consonant voicing and intraoral pressure contours in women and children). *J. Acoust. Soc. Am.*, *123*(2), 1077-1088.

[7] Müller, E. M., & Brown, W. S. 1980. Variations in the supraglottal air pressure waveform and their articulatory interpretation. In *Speech and Language: Advances in Basic Research and Practice*, edited by N. Lass Academic, Madison, WI, Vol. 4, pp. 318–389.

[8] Perrier, P. & Fuchs, S. 2015. Motor equivalence. In *The Handbook of Speech Production,* edited by M. Redford, NY: Wiley Blackwell, pp. 225-247.

[9] Perkell, J. S., Matthies, M. L., Svirsky, M. A. & Jordan, M. I. 1993. Trading relations between tongue-body raising and lip rounding in production of the vowel /u/: A pilot 'motor equivalence' study. *J. Acoust. Soc. Am.*, 93(5), 2948-61.

[10] Proctor, M. I., Shadle, C. H., & Iskarous, K. 2010. Pharyngeal articulation in the production of voiced and voiceless fricatives). *J. Acoust. Soc. Am.*, 127(3), 1507-1518.

[11] Rosen, S. 1992. Temporal information in speech: acoustic, auditory and linguistic aspects. *Phil. Trans. Roy. Soc. B: Biological Sciences*, 336(1278), 367-373.

[12] Svirsky, M. A., Stevens, K. N., Matthies, M. L., Manzella, J., Perkell, J. S., & Wilhelms-Tricarico, R. 1997. Tongue surface displacement during bilabial stops. *J. Acoust. Soc. Am.,* 102(1), 562-571.

[13] Westbury, J. R. 1983. Enlargement of the supraglottal cavity and its relation to stop consonant voicing. *J. Acoust. Soc. Am.*, 73, 1322–1336.

[14] Zygis, M., Fuchs, S. & Koenig, L.L. 2012. Phonetic explanations for the infrequency of voiced sibilant affricates across languages. Laboratory Phonology 3(2). 299–336.

Domain-initial strengthening as reduced coarticulation.

Fanny Guitard-Ivent, Cécile Fougeron
Lab. de Phonétique et Phonologie (UMR 7018) CNRS/Sorbonne Nouvelle

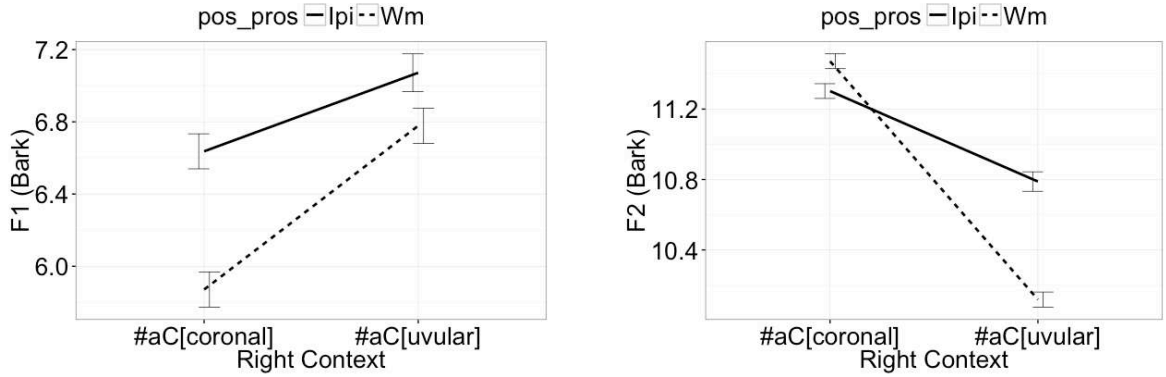
Domain-initial strengthening refers to variation in the phonetic properties of consonants and vowels when occurring after a prosodic boundary and according to the strength of this boundary (the stronger the boundary, the stronger the strengthening). This strengthening results in a *spatial* expansion of articulatory movements and the enhancement of (some of) the segment's contrastive acoustic properties (see [1], [2]). Concomitant changes in *timing* have also been reported for strengthened segments. Although lengthening of domain-initial segments has not been found to be systematic, especially in French [2, 3], overlap between consonants in #CC clusters (with "#" representing a prosodic boundary) tends to decrease with increased boundary strength [4, 5]. A recent study on French V-to-V coarticulation in $\#V_1CV_2$ sequences also showed a reduced anticipatory coarticulation of V_2 on V_1 when it is initial in a strong prosodic domain [6]. In the present study we test for a change in C-to-V coarticulation in French according to the prosodic position of the sequence (IP initial *vs.* word medial) and we look at different types of coarticulation varying in terms of cohesiveness by comparing: *carryover coarticulation* in CV sequences (where the overlap between segments is rather stable) and *anticipatory coarticulation* VC sequences (more variable and possibly more affected by other factors, as prosodic position).

About 17,000 vowels extracted from two large corpora (ESTER [7] and NCCFr [8]) of natural speech are studied. These includes /i, e, a, ɔ/ in uvular /R/ *vs.* coronal /t, d, s, z, n, l/ contexts in VC sequences. CV sequences in the same uvular *vs.* coronal contexts could also been observed for /a/ (but not for the other vowels due to insufficient number of cases). C-to-V coarticulation is compared between sequences that are either initial in an Intonational Phrase (IPi) or medial in a word (Wm). In Wm position, the opposite context (left in VC and right in CV) is a labial consonant. While in the CV sequence, C and V are always tautosyllabic, we could not control for the syllabic affiliation of C in the VC cases. Coarticulation is measured as the difference in F1 and F2 according to consonantal context. Analysis is performed using a linear mixed model in R [9] and the package lme4 [10].

Results show a clear effect of consonantal context in both VC and CV structures: as expected, F1 is significantly lower and F2 higher in coronal context than in uvular context. Prosodic position affects both F1 and F2 dimensions in a vowel-dependent way. When in IPi position, /a/ has a higher F1 (and F2 in VC only), /e/ has a lower F1 and a higher F2, /i/ has a higher F2, while no change is found for /ɔ/. More interestingly, our study reveals an interaction between the prosodic position and consonantal context. For all vowels in VC sequences, and for both F1 and F2, the effect of consonantal context is smaller in IPi position than in Wm position. This illustrated in Figure 1 for /aC/ and /eC/. In CV, the effect of consonantal context also interacts with prosodic position, although difference between prosodic positions is subtler.

This study based on a large amount of data produced in a naturalistic context shows that vowels in strong prosodic positions are less coarticulated, i.e. less overlapped by surrounding consonants, in French. These results will be discussed according to the proposed domain of activation of a π-gesture responsible of the slowing down of articulatory gestures near phrase boundaries. Overall, this reduced overlap undergone by prosodic domain initial vowels, and vowels in domain initial syllables, provides further support to the view that initial strengthening contributes to the preservation of segmental identity.

/aC/



/eC/

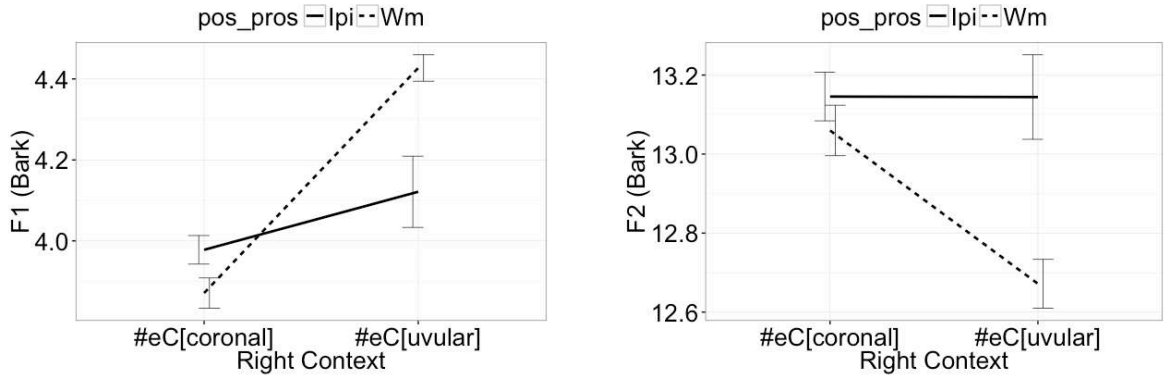


Figure 1. *F1 (left pannel) and F2 (right pannel) in Bark (as fitted by the model) for /a/ (top) and /e/ (bottom) in VC in terms of right context (coronal / uvular) and prosodic position (IPi (red) Wm (blue).*

[1] Cho, T. 2011. Laboratory phonology. *The continuum companion to phonology,* 343-368.

[2] Georgeton, L., & Fougeron, C. 2014. Domain-initial strengthening on French vowels and phonological contrasts: evidence from lip articulation and spectral variation. *Journal of Phonetics,* 44, 83-95.

[3] Keating, P., Cho, T., Fougeron, C., & Hsu, C. 2003. Domain-initial articulatory strengthening in four languages. *Laboratory Phonology,* 6*,* 143–161.

[4] Bombien, L., Mooshammer, C., Hoole, P., Rathcke, T., & Kühnert, B. 2007. Articulatory strengthening in initial German /kl/ clusters under prosodic variation. *International Congress of Phonetic Sciences XVI,* (Saarbrücken), 457–460.

[5] Byrd, D., & Choi, S. 2010. At the juncture of prosody, phonology, and phonetics—The interaction of phrasal and syllable structure in shaping the timing of consonant gestures. *Laboratory Phonology,* 10*,* 31–59.

[6] Turco, G., Fougeron, C., & Audibert, N. 2016. The effects of prosody on French V-to-V coarticulation: A corpus-based study. *Proceedings of Interspeech,* (San Francisco), 998-1001.

[7] Gravier, G., Bonastre, J-F., Geoffrois, E., Galliano, S., Tait, K. Mc., & Choukri, K. 2006. Corpus description of the ESTER evaluation campaign for the rich transcription of French broadcast news. *Proceedings of European Conference on Speech Communication and Technology,* 139–142.

[8] Torreira, F., Adda-Decker, M., & Ernestus, M. 2010. The Nijmegen corpus of casual French. *Speech Communication*, 52, 201–212.

[9] R Development Core Team. 2008. *R: A language and environment for statistical computing. R Foundation for Statistical Computing.* Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

[10] Bates, D., Maechler, M., Bolker, B., & Walker, S. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48.

Aspiration, laryngeal features, and obstruent effects on F0: Evidence from Zurich Swiss German

D. Robert Ladd[1] and Stephan Schmid[2]
University of Edinburgh[1] and *University of Zurich*[2]

Background I. In most Swiss German (SwGer) varieties, the plosives conventionally written /p b t d k g/ are *voiceless* and *unaspirated*; the main acoustic cue to the distinctions /p-b/, /t-d/, and /k-g/ is *closure duration* [2]. Phonetic accounts (e.g. [9]) often distinguish between 'fortis' /p t k/ and 'lenis' /b d g/; current phonological descriptions (e.g. [10]) often analyse 'fortis' as the geminate version of a singleton voiceless plosive. However, in many names and loanwords and some ordinary native words, /p t/ are produced with aspiration (/k/ is subject to affrication and is not considered further here). This phenomenon was first reported in the 19th century and is now spreading to more lexical items ([12]). The presence or absence of aspiration may be described as a lexically conditioned 'marginal contrast' ([4]).

Background II. Voicing contrasts in many languages are accompanied by differing effects on the fundamental frequency (F0) of a following vowel: F0 is higher after voiceless consonants than after voiced (e.g. [6]). By comparing to a nasal reference contour, [5] showed that the difference involves *raised* F0 after *voiceless* obstruents rather than lowered F0 after voiced. The difference is found regardless of the phonetic basis of the contrast (e.g. with or without prevoicing, with or without aspiration [8]). However, there is mixed evidence on the F0 effects of aspirated plosives (e.g. [3]) and on what happens in languages with 3-way laryngeal contrasts like Korean or Wu Chinese (e.g. [1]). There is also discussion of whether the differences are automatic consequences of voicing-related articulatory gestures (e.g. [11]) or intended enhancements of voicing-related phonological features (e.g. [7]).

Our study: With fortis and lenis voiceless unaspirated plosives and an apparently distinct set of aspirated plosives, SwGer can shed light on these issues. We recorded 20 university-aged female native speakers of Zurich SwGer (attached figures are preliminary, based on data from 6 speakers) reading alternative questions (like *Was it an A or a B?*). The test word was always in the *A* position to ensure a rising pitch accent; 20 test words each began with lenis, fortis, and aspirated plosives, and 20 with nasals to provide a reference contour. These were preceded by 20 practice sentences.

Findings: ***(i)*** VOT and closure duration data in Fig. 1 confirm that there are 3 phonetically distinct plosive types. ***(ii)*** F0 data in Fig. 2 show: ***(a)*** after both fortis and lenis, F0 begins higher than after nasals, and drops quite rapidly (40-60ms). The lenis contour then reconverges with the nasal contour, but after fortis, F0 begins higher than after lenis and stays higher for roughly half the vowel duration. ***(b)*** After aspirated stops, F0 begins even higher than after fortis, but typically (though there is individual variation) does *not* drop back rapidly; as with fortis, F0 remains higher for roughly half the vowel duration. ***(c) Summary:*** Aspirated plosives are distinct from both lenis and fortis in the *shape* of the F0 effect; lenis plosives are distinct from both fortis and aspirated as regards its *temporal extent.* These similarities and differences could reflect the interaction of voicing-related articulatory gestures, but may also be consistent with enhancement accounts like [7].

Conclusion: Aspiration is often seen simply as one end of a VOT continuum, with prevoicing at the other end. The SwGer evidence suggests that this is not necessarily the best way to look at things; laryngeal articulations may involve more complex interactions than is suggested by a simple distinction between 'aspirating' and 'true voicing' languages. A more nuanced understanding of the phonetic detail will shed light on proposed universal laryngeal features in phonology, and may help explain conflicting findings on obstruent F0 effects with aspirated stops in different languages.

References

[1] CHEN, Y. (2011). How does phonology guide phonetics in segment–f0 interaction? *J. Phon.* 39, 612–625.

[2] FLEISCHER, J. & SCHMID, S. (2006). Zurich German. *JIPA* 36, 243-253.

[3] FRANCIS, A. L., CIOCCA, V., WONG, V. K. M., & CHAN, J. K. L. (2006). Is fundamental frequency a cue to aspiration in initial stops? *JASA* 120, 2884–2895.

[4] HALL, K.C. (2013). A typology of intermediate phonological relationships. *The Linguistic Review* 30, 215–275.

[5] HANSON, H. M. (2009). Effects of obstruent consonants on fundamental frequency at vowel onset in English. *JASA* 125, 425–441.

[6] HOMBERT, J.-M. (1978). Consonant types, vowel quality, and tone. In *Tone: A Linguistic Survey,* ed. V. Fromkin (Academic Press, San Diego, CA), pp. 77–111.

[7] KINGSTON, J. & DIEHL, R. (1994). Phonetic knowledge. *Language* 70, 419–454.

[8] KIRBY, J. P. & LADD, D. R. (2016). Effects of obstruent voicing on vowel F0: Evidence from "true voicing" languages. *JASA* 140, 2400-2411.

[9] KOHLER, K.J. (1984). Phonetic explanation in phonology: the feature fortis/lenis. *Phonetica* 41, 150–174.

[10] KRAEHENMANN, A. (2001). Swiss German stops: geminates all over the word. *Phonology* 18, 109–145.

[11] LÖFQVIST, A., BAER, T., MCGARR, N. S., & STORY, R. S. (1989). The cricothyroid muscle in voicing control. *JASA* 85, 1314–1321.

[12] SCHIFFERLE, H.-P. (2010). Zunehmende Behauchung: Aspirierte Plosive im modernen Schweizerdeutsch. In *Alemannische Dialektologie: Wege in die Zukunft.* (*Zeitschrift für Dialektologie und Linguistik, Beiheft 141*), pp. 43-55.

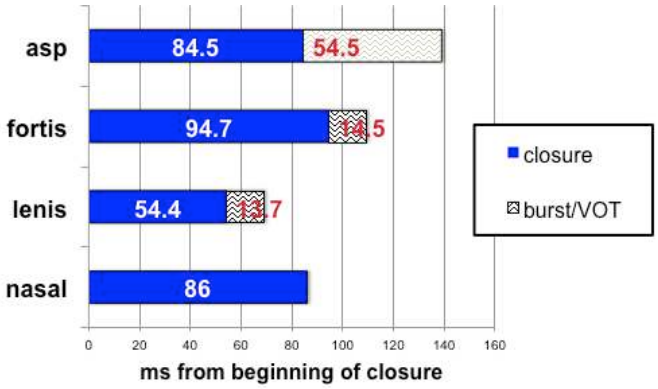


Figure 1. Mean closure duration and VOT for the 3 types of stop and the nasal.

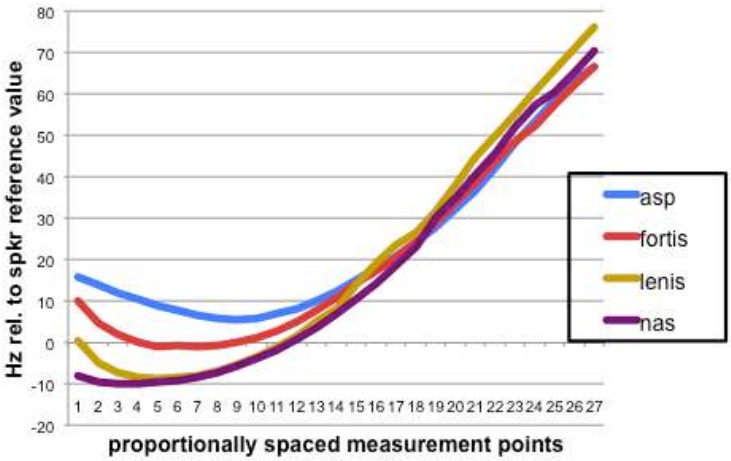


Figure 2. Mean F0 contours on the stressed syllable for the 3 types of stop and the nasal. F0 is normalised for individual differences of pitch range and normalised values are pooled across speakers.

Implicit Social Cues Influence the Interpretation of Intonation

James S. German
Aix Marseille Université, CNRS, LPL, Aix-en-Provence, France

Individuals may have significant experience with more than one variety of their native language. If the non-native varieties are saliently linked to specific social identities, then an individual's production or perception can be biased towards a particular variety through contextual cues that exemplify the associated identity (e.g., [1], [2]). In this study, we explore whether the interpretation of intonational patterns by Singapore English (SgE) listeners is influenced by implicit social cues. Specifically, we ask whether the relationship between accentuation and pronoun reference, which is robust in American (and other) varieties, but weak in SgE, varies with exposure to either a "Singaporean" or "American" contextual cue.

It is well-documented that the reference of personal pronouns in many English varieties depends on intonation, and specifically, whether a pronoun bears a pitch accent or not. Consider the following sentence: (1) *Tim laughed at Shawn, then Stacy laughed at him*. American English (AmE) listeners generally interpret unaccented *him* as referring to *Shawn*, and accented *him* to *Tim*. While these effects have been shown to depend on focus presuppositions associated with clause-level intonation patterns (e.g., [3]), what is of interest to our study is the role of prominence. The realization of this contrast depends on the fact that AmE freely places pitch accents in service of information structure. By comparison, the intonation system of SgE is edge-based: it does not have pitch accents and prominence is determined primarily by prosodic phrasing. Specifically, each Accentual Phrase (AP) begins with a L tone at its left edge, a H tone at its right edge, and has significant lengthening on the final syllable [4]. As a consequence, the prominence of a pronoun depends on its position in a phrase, and does not correlate with information structure. Object pronouns tend to fall at the right edge of an AP and generally receive high prominence (which is likely responsible for the impression that SgE speakers stress pronouns inappropriately). Together, these facts predict that SgE listeners are insensitive to prominence in computing pronoun reference as compared with AmE listeners, which we confirmed in a pilot study. Nevertheless, most SgE individuals have substantial contact with AmE, suggesting that their system may adapt based on the regional identity of the speaker and other cues.

In our study, 40 SgE listeners were presented with spoken sentences, which varied in the accentual status of the object pronoun (as in (1)). Participants responded by choosing from two paraphrases, which reflected reference to either the subject or the object of the previous clause. In addition, participants were exposed before and during the experiment to either a "Singaporean" cue or an "American" cue, which took the form of an image of either a popular Singaporean or American television series. The image was construed as being present by accident and was therefore not explicitly linked to the identity of the speaker. The speaker was a native Singaporean with training which allowed him to produce AmE intonational contrasts while maintaining SgE segmental features. Our hypothesis was that if the contextual cues bias the listeners' toward specific systems, then they should show more sensitivity to accentual status in the American condition than in the Singaporean condition.

The results confirmed our hypothesis. Specifically, there was a significant interaction between cue type and accentual status in the expected direction. Contrary to pilot results, however, sensitivity to accentual status was not correlated with listeners' self-reported exposure to AmE. Our findings point towards an exemplar basis for the representation of intonation, in line with findings from explicit training [5] and collocational frequency [6], though this is the first study that shows socio-indexical effects at the intonation-meaning interface. Future studies will explore at which level of representation the socio-indexical cues enter into the overall mapping from phonetic inputs to discourse-level meaning.

References

[1] Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of language and social psychology, 18(1)*, 62-85.

[2] Hay, J., Warren, P., & Drager, K. (2006). Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics, 34(4)*, 458-484.

[3] McClay, E. & Wagner, M. (2014). Accented Pronouns and Contrast. *Proceedings of the 50th Meeting of the Chicago Linguistics Society in 2014*.

[4] Chong, A. J. & German, J. S. (to appear). The accentual phrase in Singapore English. *Phonetica*, 74(2), 63-80.

[5] Kurumada, C., Brown, M., Bibyk, S., Pontillo, D. & Tanenhaus, M. (2014) Is it or isn't it: Listeners make rapid use of prosody to infer speaker meanings. *Cognition, 133(2)*, 335-342.

[6] Calhoun, S. & Schweitzer, A.. (2012) Can intonation contours be lexicalised? Implications for discourse meanings. *Prosody and Meaning (Trends in Linguistics)*, *25*, 271-327.

Who likes liver? How German speakers use prosody to mark questions as rhetorical

Jana Neitsch, Daniela Wochner, Katharina Zahner, Nicole Dehé, Bettina Braun
University of Konstanz

While an *information-seeking question* (ISQ) elicits information [1, 2], the implied answer to a *rhetorical question* (RQ) is usually part of the common ground (e.g. [1, 3]). RQs have as yet mainly been the subject of pragmatic and semantic investigations, but research on their prosodic realisation is still rare. In a first study on the production of RQs in German, [4] found that RQs differed prosodically from ISQs in *final boundary tone*, *pitch accent type*, *duration*, *voice quality*, and *initial pitch*. Compared to [4], the present study controls the discourse status of the syntactic object, such that it is always discourse given, as well as the inherent rhetoricity between *wh*- and polar questions by keeping the contexts the same.

We constructed 22 context-question quadruplets (Tab. 1), manipulating question type (*wh* vs. polar; half of the contexts were presented with polar and half with *wh*-questions) and illocution type of context (ISQ vs. RQ; distributed within-subjects). Forty participants were randomly assigned to one of two lists and tested individually. After having read the context silently from a screen, they produced the target question as naturally as possible. So far, data of 12 participants (∅=21.7 years, SD=2.27 years, 2 male) have been analysed (n=512, *wh*: 129 ISQs, 130 RQs; polar: 127 ISQs, 126 RQs). Using *Praat*, two trained annotators labelled the words (Tier 1), accented syllable(s) (Tier 4), pitch accents and boundary tones (following GToBI [5], Tier 5; see Fig. 1). In case of disagreement, a consensus label was found. We also extracted vocal effort (H1*-A3*, [6]) at three vowels of interest (Tier 3, Fig. 1).

Boundary tones: For polar questions, RQs were predominantly realised with a high plateau (H-%: 74%), while ISQs were mostly produced with a high rise (H-^H%: 87%). For *wh*-questions, RQs mostly show low boundary tones (L-%: 94%), while ISQs were more variable: high rises (33%), low rises (19%) and low boundary tones (45%); Tab. 2.

Nuclear pitch accents: For polar questions, both ISQs and RQs were most often realised with a low-pitched accent (L*, ISQ: 84%, RQ: 77%), followed by a rise (high rise or high plateau). For *wh*-questions, there is a difference in nuclear pitch accents, however. RQs show a higher proportion of L*+H nuclear accents (55%) than ISQs (3%), while ISQs show a higher proportion of L+H* nuclear accents (43%) than RQs (22%). Moreover, in ISQs, L* (28%) and H+!H* (15%) are more common than in RQs (L*: 9%, H+!H*: 2%); Tab. 2.

Acoustic measurements: Tab. 2 summarizes the means of RQs and ISQs in the acoustic variables *initial pitch*, *duration*, and *vocal effort*, listed by question type (p-values in brackets; note that the significance levels were corrected according to [7]). Our findings show that RQs, compared to ISQs, have a longer utterance duration (for both *wh*- and polar questions), longer normalised durations of the first word (for *wh*-questions only) and the final object (for both *wh*- and polar questions); RQs show a tendency for lower initial pitch (for *wh*-questions only), and breathier voice quality in all measured vowels (for *wh*-questions only).

These results are comparable to [4] (except for *voice quality* and *initial pitch*), corroborating [4]'s findings in two respects: First, our results suggest that the observed differences can be attributed to illocution type and are not merely an artefact of information-structure. Second, our results support [4]'s findings that the differences between illocution types are realised more clearly in *wh*-questions than in polar questions. Thus, speakers provide clear prosodic cues to mark interrogatives as rhetorical, at least in *wh*-questions. Likewise, in a recent perception study (investigating *accent type* and *voice quality*), we show that these cues are used by listeners in order to identify RQs. We are currently analysing more data to shed more light on the reported findings regarding *initial pitch* and *voice quality* on the one hand, and the difference in the realisation of rhetoricity across question types on the other. We also focus on the production-perception link, testing cues, which have been identified as markers of rhetoricity in production, also in perception.

	ISQ	RQ
polar	"You serve your relatives a meat platter with sauerkraut, liver and many other things. You want to know whether someone would like to have some of it or not." ***Mag denn jemand Leber?*** *"Does anyone like liver?"*	"During coffee your aunt offers your relatives a tray with liver. However, obviously no one wants this, because it does not go with coffee and cake." ***Mag denn jemand Leber?*** *"Does anyone like liver?"*
wh	"You serve your relatives a meat platter with sauerkraut, liver and many other things. You want to know which of them would like to have some of it." ***Wer mag denn Leber?*** *"Who likes liver?"*	"During coffee your aunt offers your relatives a tray with liver. However, obviously no one wants this, because it does not go with coffee and cake." ***Wer mag denn Leber?*** *"Who likes liver?"*

Tab. 1: *Example context-question quadruplet in an information-seeking and rhetorical context. Wh-questions always contained the wh-element 'who' and polar questions always included 'anyone'.*

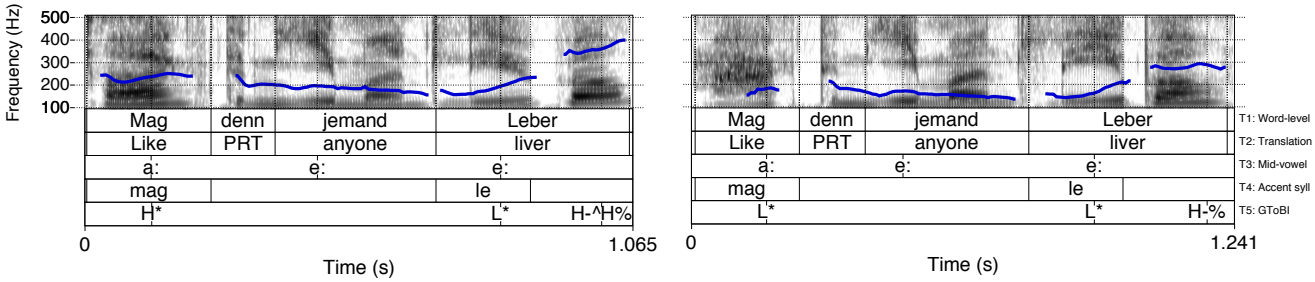


Fig. 1: *Example polar question, ISQ (left) and RQ (right), showing the annotation layers.*

GToBI labels		**polar questions**	***wh*-questions**
Final boundary:	H-^H%	RQ: 21% vs. ISQ: 87%	RQ: 4% vs. ISQ: 33%
	H-%	RQ: 74% vs. ISQ: 4%	RQ: 2% vs. ISQ: 2%
	L-%	RQ: 4% vs. ISQ: 4%	RQ: 94% vs. ISQ: 45%
Nuclear accent:	L*+H	RQ: 3% vs. ISQ: 0%	RQ: 55% vs. ISQ: 3%
	L+H*	RQ: 6% vs. ISQ: 9%	RQ: 22% vs. ISQ: 43%
	L*	RQ: 77% vs. ISQ: 84%	RQ: 9% vs. ISQ: 28%
	H+!H*	RQ: 0% vs. ISQ: 0%	RQ: 2% vs. ISQ: 15%
Acoustic measurements		**polar questions**	***wh*-questions**
Initial pitch		RQ: 213Hz vs. ISQ: 224Hz (p=0.16)	RQ: 207Hz vs. ISQ: 216Hz (p=0.08)
Utterance duration		RQ: 1397ms vs. ISQ: 1207ms (p<0.0001)	RQ: 1288ms vs. ISQ: 1085ms (p<0.0001)
Norm. duration (final object) [4]		RQ: 47.6% vs. ISQ: 46.2% (p<0.03)	RQ: 54.7% vs. ISQ: 52.3% (p<0.02)
Norm. duration (first word) [4]		RQ: 18.0% vs. ISQ: 17.4% (p=0.16)	RQ: 11.0% vs. ISQ: 11.7% (p<0.02)
Vocal effort (H1-A3*)*		vowel 1, RQ: 19.6dB vs. ISQ: 18.7dB (p=0.90) vowel 2, RQ: 19.5dB vs. ISQ: 20.2dB (p=0.14) vowel 3, RQ: 21.2dB vs. ISQ: 20.0dB (p=0.22)	vowel 1, RQ: 16.1dB vs. ISQ: 14.3dB (p<0.05) vowel 2, RQ: 17.6dB vs. ISQ: 17.3dB (p<0.05) vowel 3, RQ: 22.1dB vs. ISQ: 22.0dB (p<0.0001)

Tab. 2: *Mean frequency of occurrence of GToBI labels for RQs and ISQs and means for acoustic measurements (RQs and ISQs, Benjamini-Hochberg corrected significance level).*

[1] Groenendijk, J. & M. Stokhof. 1984. *Studies on the semantics of questions and the pragmatics of answers.* Amsterdam: Universiteit van Amsterdam, PhD Thesis.

[2] Meibauer, J. 1986. *Rhetorische Fragen* (Linguistische Arbeiten 167). Berlin: De Gruyter.

[3] Caponigro, I. & J. Sprouse. 2007. Rhetorical questions as questions. *Proc. of Sinn und Bedeutung 11*, 121–133.

[4] Wochner, D., J. Schlegel, N. Dehé & B. Braun. 2015. The prosodic marking of rhetorical questions in German. *Proc. of Interspeech* 2015, 987–991.

[5] Grice, M., S. Baumann, & R. Benzmüller. 2005. German intonation in autosegmental-metrical phonology. In J. Sun-Ah (Ed.). *Prosodic Typology.* Oxford: OUP, 55-83.

[6] Mooshammer, C. 2010. Acoustic and laryngographic measures of the laryngeal reflexes of linguistic prominence and vocal effort in German. *JASA, 127*(2). 1047-1058.

[7] Benjamini, Y. & Y. Hochberg. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological), 57*(1). 289-300.

Flexibility in the association of tones captures melodic alternations in Spanish

Francisco Torreira[1,2] & Martine Grice[3]
[1]*McGill University* & [2]*Max Planck Institute for Psycholinguistics*
[3]*University of Cologne*

This paper explores the structure of intonation contours, arguing for the existence of melodic constructions that cannot be analyzed as simple sequences of the types of units traditionally used in phonological models of intonation (i.e. pitch accents and edge tones; prenuclear and nuclear contours). Following a discussion of melodic constructions in English [1, 2], Catalan [3], and Spanish [4], we provide data from two imitation-and-completion experiments, each carried out on a different Spanish melodic construction (the low-rise-fall and the circumflex contour), and involving five native speakers of Castilian Spanish, and five Italian L2 learners of Spanish of different proficiency levels. Both experiments employed an initial training phase requiring participants to imitate stimuli of fixed length and always exhibiting the same pitch contour in a constant pragmatic/focus context. The subsequent test phase required participants to produce new utterances of one, two and three prosodic words but still in the same pragmatic/focus context (5 items * 3 phrase lengths * 5 speakers * 2 native languages = 150 trials per experiment).

We show that, for both of the studied melodic constructions, the nuclear contours of native speakers alternate as a function of the number of prosodic words in the phrase. In the native speakers' low-rise-falls, the nuclear contour in *El hermano de Ma**no**lo* (Manolo's brother; nuclear syllable in bold face) has a high pitch accent and a low edge tone, whereas in shorter phrases, such as *Ma**no**lo*, the nuclear accent is low, and a high target occurs as an edge tone on the final syllable (Figure 1). In circumflex contours produced with phrases of two or three prosodic words, native speakers exhibited nuclear contours consisting of an upstepped high pitch accent and a low edge tone, whereas in short phrases of one prosodic word, their nuclear contours were consistently rising (a high pitch accent followed by an even higher boundary tone; Figure 2). L2 learners of Spanish, on the other hand, generally produced similar nuclear contours (low-rises and upstepped rise-falls) regardless of phrase length.

To account for these melodic alternations in the native speakers' productions, we will argue for a clearer separation between tones and metrical structure in intonational phonology, allowing melodic constructions in the intonational lexicon-grammar of a language to have tones without an intrinsic culminative function (as pitch accents) or delimitative function (as edge tones). Instead, tones are associated in a context-dependent way, becoming either pitch accents or edge tones depending on available docking sites. More generally, our data support the existence of meaningful intonational units exhibiting more flexibility in the association properties of tones than traditionally discussed in the intonational phonology literature.

References

[1] Sag, I. A., & Liberman, M. (1975). The intonational disambiguation of indirect speech acts (Vol. 11, pp. 487–497). *Proceedings of the Chicago Linguistic Society*.

[2] Grice, M. (1995). Leading tones and downstep in English. *Phonology*, 12(02), 183–233.

[3] Prieto, P. (2002). Tune-text association patterns in Catalan: an argument for a hierarchical structure of tunes. *Probus*, 14, 173–204.

[4] Torreira, F. (2015). Melodic alternations in Spanish. *Proceedings of ICPhS 2015*.

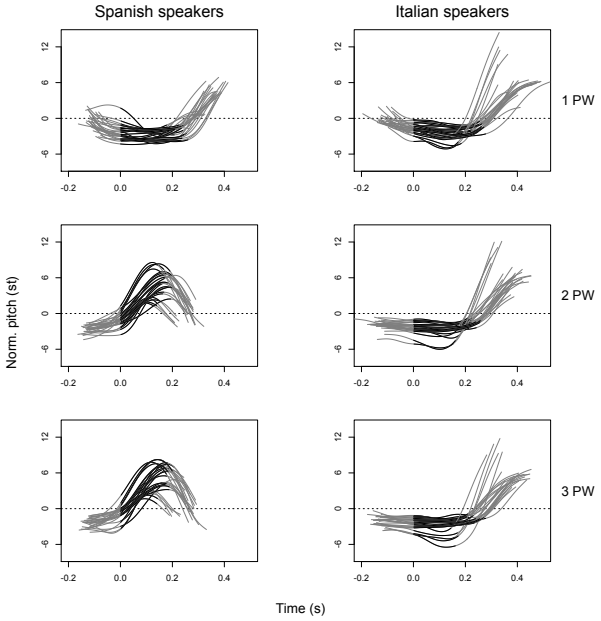


Figure 1 Low-rise-falls: pitch contours (in semitones, relative to each speaker's median) produced over the last prosodic word of each target utterance (*Ma**no**lo* in all cases) pooled by language group and phrase length (# of prosodic words). The portion of each pitch contour corresponding to the syllable with nuclear stress (*no* in all cases) is drawn in black.

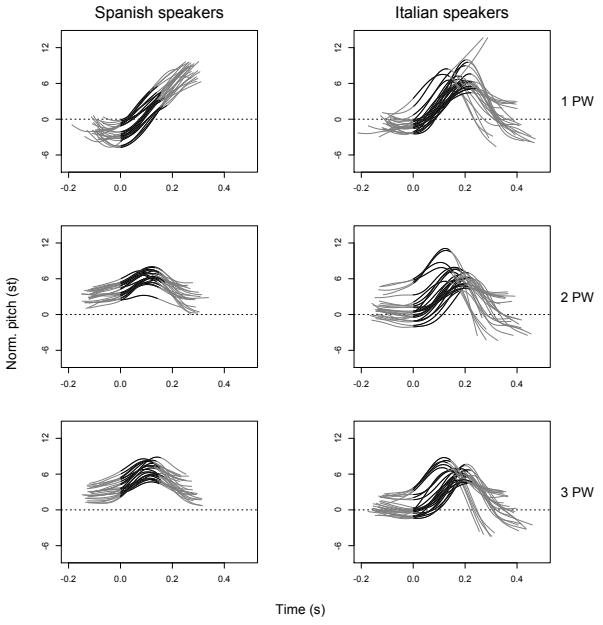


Figure 2 Circumflex contours: pitch contours (in semitones, relative to each speaker's median) produced over the last word of each target utterance (*Ma**no**lo* in all cases) pooled by language group and phrase length (# of prosodic words). The portion of each pitch contour corresponding to the syllable with nuclear stress (*no* in all cases) is drawn in black.

Interpreting Rhetorical Questions: the Influence of Pitch Accent Type, Voice Quality and the Modal Particle *denn*

Marianne Kusterer, Jana Neitsch, Bettina Braun, Nicole Dehé
University of Konstanz

While *information-seeking questions* (ISQs) function to elicit information [1, 2], the implied answer to a *rhetorical question* (RQ) is usually assumed to be already known to all interlocutors [3]. In German, modal particles are among the formal linguistic devices signalling RQs (e.g. *schon, auch* [2]). For *wh*-questions, the particle *denn* has been argued to be almost obligatory as a question marker [4]. However, its influence on question interpretation (RQ or ISQ) has not yet been tested empirically. Furthermore, little is known about the prosody of RQs. In a first systematic comparison of the prosodic realisation of string-identical RQs and ISQs, [5] report, inter alia, a breathier voice quality (VQ) for RQs and more L*+H accents in *wh*-RQs than in ISQs. The current work focuses on the perceptual relevance of VQ, pitch accent type and the presence vs. absence of the particle *denn* for the identification of *wh*-questions as RQs or ISQs, tested in two forced-choice perception experiments.

In Experiment 1, *VQ* (breathy vs. modal) and *nuclear pitch accent type* (late peak accent L*+H vs. early peak accent H+!H*) were manipulated within-subjects in 32 string-identical *wh*-questions containing *denn* (RQ vs. ISQ reading, e.g. *Wer mag denn Vanille,* 'Who likes PRT vanilla'). Each *wh*-question was produced and recorded in four experimental conditions (crossing *pitch accent type* and *VQ*) by a female native speaker of German, resulting in 128 experimental items (see Figure 1 for example contours). Word durations were manipulated so that they did not differ between VQ conditions. For Experiment 2, the particle was cut out of the recordings without effecting the naturalness of the materials.

Twenty-four native speakers of German participated in each experiment. In each trial, they were first visually presented with a picture of the syntactic object of the target interrogative to provide a non-linguistic contextual setting. After 2500ms, the two labels 'real question' and 'rhetorical question' (representing ISQ and RQ respectively) were shown in the top corners of the screen (position of the boxes was counterbalanced across conditions). Another 1000ms later, participants heard the *wh*-question in one of the four conditions over headphones and had to decide on the corresponding label by mouse click. Clicks were analysed for each experiment using logistic mixed-effects models with *VQ* and *accent type* as fixed factors and *participants* and *items* as crossed random factors. For both experiments (with and without *denn*), we found significant effects of *accent type* and *VQ* (see Figure 2). Participants clicked on RQ more often when the target was realised with a late peak accent (both experiments $p<0.01$) and when it was produced with breathy VQ (both $p<0.01$). Thus, most clicks on RQ occurred when the *wh*-interrogative was produced with breathy VQ and a late peak accent (L*+H), independent of the presence of the particle *denn*. An ISQ interpretation was most frequent for targets produced with modal VQ and an early peak accent (H+!H*). Furthermore, there was a significant interaction between *accent type*, *voice quality* and *particle* ($p<0.001$). Hence, the presence of the particle strengthened the respective interpretations for both RQ and ISQ (see Figure 2).

In line with [6], we propose that the acoustic configuration *nuclear late peak accent* plus *breathy VQ* is a conventionalised prosodic contour for *wh*-questions, which facilitates listeners' disambiguation towards an RQ reading between string-identical and semantically ambiguous interrogatives. This assumption is strengthened by the results of Experiment 2, which shows that this prosodic configuration is still reliably interpreted as rhetorical in the absence of the modal particle. Given that there are differences as to how strongly an ISQ reading is favored in the early peak plus modal voice quality condition with and without *denn,* a third experiment will combine the two experiments in a within-subjects design. The data gathered by this future experiment will allow further claims about the extent to which the German modal particle *denn* interacts with prosodic parameters.

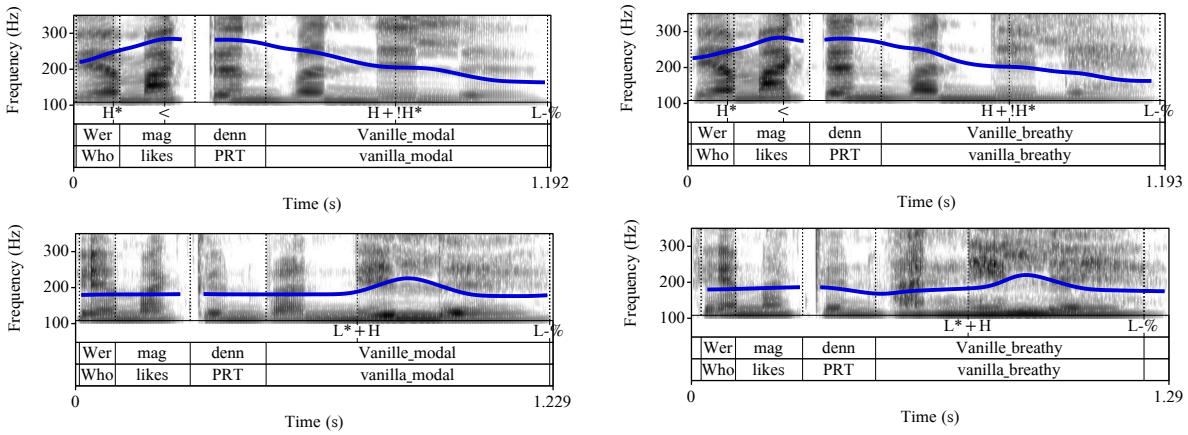


Figure 1. *Example wh-question produced in all four experimental conditions (left panel displays productions in modal voice).*

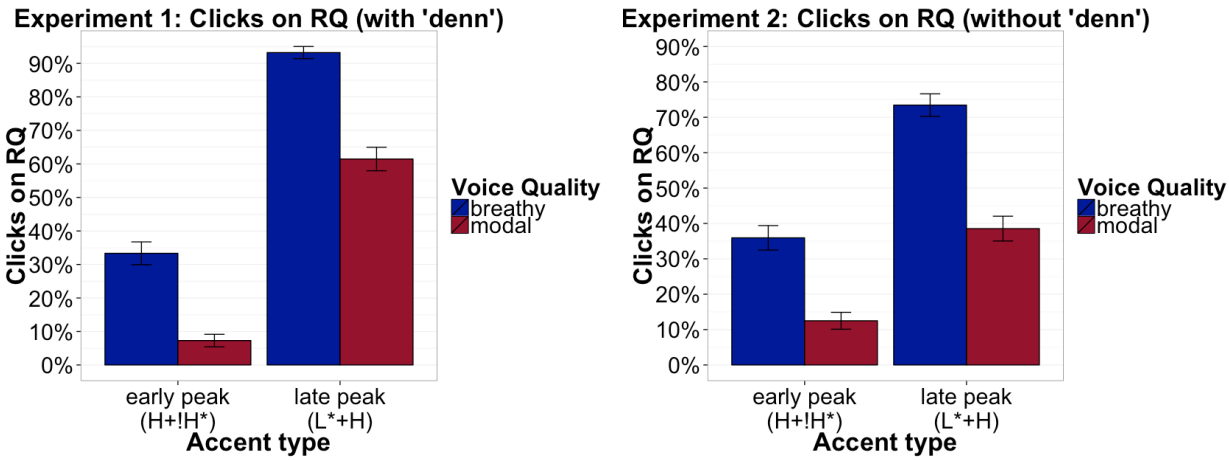


Figure 2. *Clicks on RQs by accent type (early peak vs. late peak) and voice quality. Whiskers indicate standard errors.*

References

[1] Groenendijk, J. & Stokhof, M. 1984. *Studies on the semantics of questions and the pragmatics of answers.* Amsterdam: Universiteit van Amsterdam, PhD Thesis.

[2] Meibauer, J. 1986. *Rhetorische Fragen* (Linguistische Arbeiten 167). Berlin: De Gruyter.

[3] Caponigro, I. & Sprouse, J. 2007. Rhetorical questions as questions. *Proc. of Sinn und Bedeutung* 11, 121–133.

[4] Thurmair, M. 1991. *Zum Gebrauch der Modalpartikel 'denn' in Fragesätzen: Eine korpusbasierte Untersuchung* in E. Klein, F. Pouradier Duteil & K. H. Wagner (eds.), *Betriebslinguistik und Linguistikbetrieb;* Linguistische Arbeiten 260, Tübingen: Niemeyer, 377–388.

[5] Wochner, D., Schlegel, J., Dehé, N., & Braun, B. 2015. The prosodic marking of rhetorical questions in German. *Proceedings of INTERSPEECH 2015*, 987–991.

[6] Hellbernd, N., & Sammler, D. 2016. Prosody conveys speaker's intentions: Acoustic cues for speech act perception. *Journal of Memory and Language* 88, 70–86.

From lab speech to dialogue corpora:
***Givenness* and *Obviousness* in Castilian Spanish intonation**

Jan Fliessbach
Freie Universität Berlin

Based on data from *Discourse Completion Tasks* and other (semi)controlled forms of laboratory speech, intonation research currently assumes a large set of pragmatic meanings to be encoded phonologically in Castilian Spanish declarative sentences. One of them is *Obviousness*, a grammatical category that can be defined as a match between the proposition of a sentence and the expectations inferred from the *Common Ground* of the interlocutors. It is different from *Givenness* in that it operates on the entire proposition, not only on those parts of a sentence with a close antecedent [1]. In Castilian Spanish, *Statements of the Obvious* are assumed to be coded by a L+H* L!H% nuclear intonation contour [2]. Moreover, it has been claimed that information that is already known can be marked by a continuously rising F0 until an intermediate boundary tone (H-), which in turn marks the end of the information already available to the interlocutors [3]. It remains an open question whether this configuration, which is similar (though not identical) to what has been described as a *Continuation rise* [4], should be interpreted as a marking of *(inferentially) given* or *obvious* information. Though detailed and phonologically predictive, these findings share the common problem that they are based on a relatively small number of controlled speech productions. Hence, little is known to date about the generalizability of such observations to spontaneous speech and about the phonetic implementation of these categories. In a nutshell, it remains largely unknown if such intonational indicators of different semantic and pragmatic subtleties can be shown to combine freely in spoken dialogue.

In intonation research, the step from laboratory speech to spontaneous corpus data is notoriously difficult, partly due to the lack of prosodic annotation in large oral corpora. Without the possibility to search automatically for specific tonal configurations, natural occurrences of rare intonational phenomena become needles in a costly and time-consuming haystack. The present study is an attempt to sidestep this problem by searching for simultaneous expression of modal meaning through lexical and intonational means. We conducted search queries for particles such as *pues* 'so/well', *hombre* 'man', *tío* 'uncle', *tía* 'aunt', *claro* 'clear', *obvio* 'obvious' and *por supuesto* 'of course' in a spontaneous oral dialogue sub-corpus of Castilian Spanish of approximately 100.000 words [5], and are currently applying the same method to a comparable corpus with higher recording quality of approximately 150.000 words [6]. We argue that these particles, if used utterance-initially, can denote *Obviousness* [7] and ask which intonational contours occur in the utterances they are associated with. Our results from corpus [5] show that the rising contour ending in a high boundary tone (H-) accompanies these modal particles in almost all sentences (for which $N_{Verb} \geq 1$) and falls on the inflected verb of the asserted part of the utterance in most cases. In the final (nuclear) part of the respective utterances, we find combinations of L+H* with simple boundary tones (H% or L%), rather than the complex L+H* L!H% nuclear contour discussed in the standard literature on intonation in Spanish (see example (1) and figure 1).

Unlike previous research, the present study yields insights on modal and information-structural meaning encoded in Castilian Spanish spontaneous conversation. It adds to the existing evidence for the importance of prenuclear intonation for the pragmatic interpretation of utterances. And it presents an innovative lexical-query method for accessing rare intonational configurations in oral corpora.

(1) ˈombɾe maˈma lo ˈtjenen ke paˈɣaɾ poɾ ˈkwaɲðo se an meˈtiðo en alˈɣun ˈlaðo
Man, mama, they have to pay it for when they have gone somewhere!

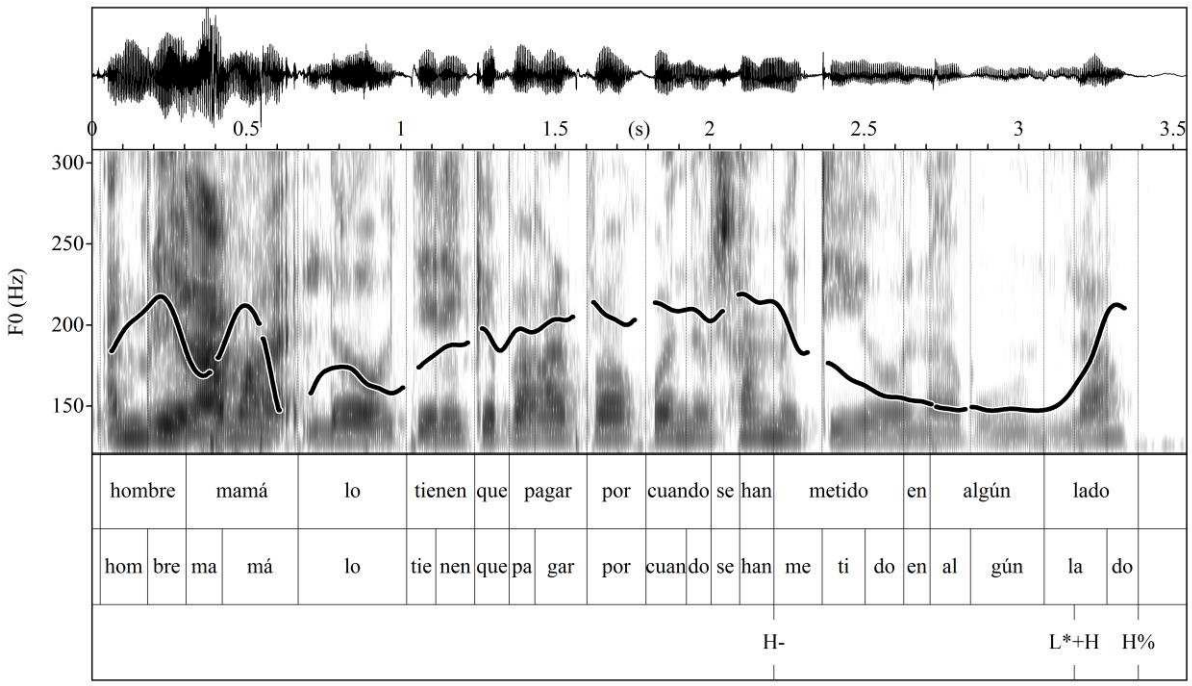


Figure 1. *Discourse particle 'hombre' preceding Castilian Spanish declarative sentence.*

[1] Riester, A., & Baumann, S. (2014). *RefLex Scheme - Annotation Guidelines*. Version of August 1, 2014. Retrieved from http://www.ims.uni-stuttgart.de/institut/mitarbeiter/arndt/doc/RefLex-guidelines-01aug-2014.pdf

[2] Hualde, J. I., & Prieto, P. (2015). Intonational variation in Spanish. In S. Frota & P. Prieto (Eds.), *Intonation in Romance* (pp. 350–391). Oxford University Press. doi:10.1093/acprof:oso/9780199685332.003.0010

[3] Hualde, J. I. (2014). *Los sonidos del español: Spanish Language edition*. Cambridge: Cambridge University Press.

[4] Feldhausen, I. (2016). Inter-speaker variation, Optimality theory, and the prosody of clitic left-dislocations in Spanish. *Probus*, *28*(2). doi:10.1515/probus-2015-0005

[5] Cresti, E., & Moneglia, M. (2005). *C-ORAL-ROM: Integrated reference corpora for spoken Romance languages. Studies in Corpus Linguistics: v. 15*. Amsterdam, Philadelphia, PA: J. Benjamins.

[6] PRESEEA. (2014-). Corpus del Proyecto para el estudio sociolingüístico del español de España y de América. Retrieved from [http://preseea.linguas.net]

[7] Briz, A., Pons Bordería, S., & Portolés, J. (Eds.). (2011). *Diccionario de partículas discursivas del español*. Valencia. Retrieved from www.dpde.es

CV co-occurrence and articulatory control in three Brazilian children from 0:06 to 1:07

Mariana Hungria and Eleonora C. Albano
University of Campinas

Many studies in language acquisition have addressed consonant/vowel co-occurrence, henceforth CV. Traditional views focus on the upper vocal tract (Davis, B.L., Macneilage, P.F., 1992), while a recent one stresses the importance of its lower end (Esling, J., 2012). None has nevertheless attempted to understand how these two tracts interact and cooperate in early vocalization. Besides, the role of the ambient language in this interaction is also understudied. Although there are common preferences for certain CV combinations, sensitivity to language environment may already be present in babbling (Boisson-de-Bardies, B.; Vihman, M., 1991).

The aim of this paper is to integrate such apparently contradictory views by reporting on research that observed activity in the entire vocal tract during the emergence of CV combinations in three Brazilian children acquiring BP in interaction with their parents, between the ages of 0:06 and 1:07. We used a mixed longitudinal and cross-sectional method, whereby one child was followed longitudinally while the other two were observed at later, complementary stages. Data were collected using a digital recorder, transcribed with the aid of acoustic analysis, and later processed with a syllable counter. Table 1 displays longitudinal data for P1 at 0:06-0:07, while table 2 shows cross-sectional data for P3 at 1:05-1:07. Regarding buccal consonants, P1 seems to prefer coronals and dorsals, whereas P3, observed at a later stage, seems to have followed a different path, favoring labials. As for vowels, P1's front vowels were expressive; on the other hand, they were the minority for P3. Back vowels are the minority for P1, but are in second position for P3, after central vowels.

Our results uncovered the following trends: i) biomechanical constraints interact with ambient language influences; ii) control over articulations seems to be different for vowels and consonants; iii) infants tend to have a favorite babbling vowel; iv) the lower vocal tract remains active past 12 months; v) not all children acquire articulatory control in the same way.

Overall, our data do not support any specific view, but, rather, call for the integration of several separate strands in the literature. In addition, they point to the need for a thorough study of the sounds of child-adult interaction, considering details of adult expressive vocalization that might elucidate the role of the environment in children's early preference for laryngeals.

		V			N
		front	back	central	
C	labial	14	6	31	51
	coronal	58	6	185	249
	dorsal	35	3	15	53
	laryngeal	43	12	280	335
Total		150	27	511	688

Table 1 -– P1's raw CV combination count (0:06-0:07)

		V			N
		front	back	central	
C	labial	0	10	58	68
	coronal	3	46	1	50
	dorsal	3	0	0	3
	laryngeal	12	38	19	69
Total		18	30	142	190

Table 2 – P3's raw CV combination count (1:05-1:07)

[1] Boisson-de-Bardies, B.; Vihman, M. 1991. Adaptation to Language: Evidence from Babbling and First Words in Four Languages**.** *Language, Vol. 67, No. 2*, pp. 297-319, 1991.
[2] Browman, C.; Goldstein, L.1992. Articulatory Phonology: an overview. *Phonetica, 49*, p. 155-180.
[3] Esling, J**.** 2012**.**The Articulatory Function of the Larynx and the Origins of Speech. *Proceedings of the Annual Meeting of the Berkeley Linguistics Society 38.*

Acoustic characteristics of novel English adjective-noun constructions

Marcel Schlechtweg
University of Kassel

The current paper aims at analyzing novel/non-lexicalized adjective-noun (AN) combinations in American English and asks whether one specific factor, namely semantic (non)-compositionality, has an influence on their stress pattern.[1] It is hypothesized that non-compositional semantics trigger a higher degree of initial stress, i.e. stress on the adjective, than compositional semantics. The second question to be addressed is whether stress is placed differently if another device that emphasizes non-compositionality, namely *called so*, is used together with non-compositional constructions.

Six native speakers of American English participated in a production study that investigated six disyllabic AN constructions and the independent variable SEMANTIC COMPOSITIONALITY (within-subject/item), which had the following three levels:

(1) Compositionality (= C): *Thomas took a black tram again, which has a color he likes.*

(2) Non-compositionality (without *called so*) (= N): *Thomas took a black tram again, which is a tram that runs only during the night.*

(3) Non-compositionality (with *called so*) (= S): *Thomas took a black tram again, which is called so because it is a tram that runs only during the night.*

In each condition, subjects (1) read the sentence silently, (2) had to answer a question referring to the sentence in order to ensure that they had understood the meaning (e.g. *Is a black tram a tram that goes to the graveyard?* (Correct answer: No)) and (3) read the sentence aloud and were recorded with Praat (Boersma & Weenink 2016). In each condition, an item occurred in the same phonetic environment as in the other conditions. The ratio and difference of the durations/intensities/F0s of the vowel of the adjective (e.g. *black*) and the vowel of the noun (e.g. *tram*) of each complex item was created in each of the three conditions. It was assumed that a greater ratio/difference signaled a higher degree of initial stress. N triggered, in terms of duration and F0, a significantly higher degree of initial stress than C and S, which, in turn, did not cause any significantly different degree of initial stress. The results show that non-compositionality triggers a higher degree of initial stress than compositionality. However, if another device that indicates non-compositionality is used (*called so*), the degree of initial stress drastically decreases because the non-compositional semantics are already signaled by means of *called so*. The results of the ratios of the durations are given as an example in Figure 1.

[1] Note that, in order to exclude the influence of lexicalization, only novel, i.e. non-lexicalized constructions were tested in the study. In this respect, the experiment differs from several others that investigated lexicalized items such as *greenhouse* (cf. McCauley, Hestvik, & Vogel 2012; Morrill 2011; Vogel & Raimy 2002).

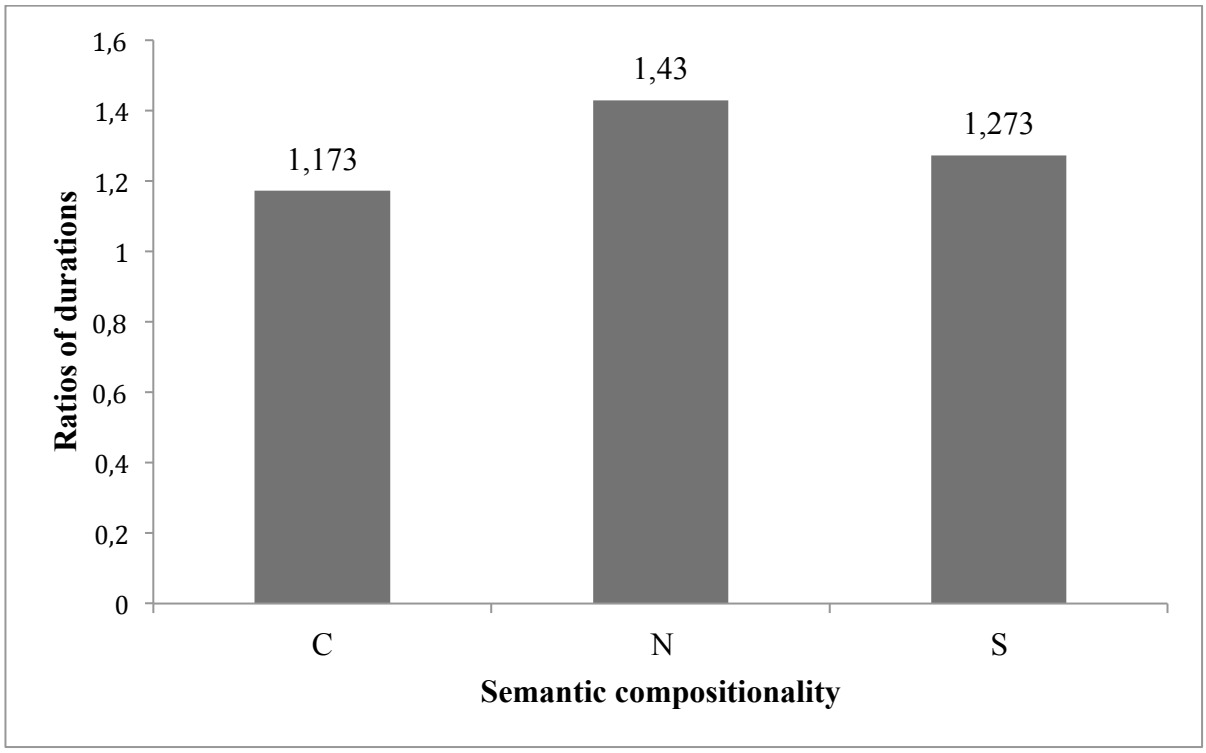


Figure 1: Ratios of the durations

References

[1] Boersma, P., & Weenink, D. 2016. *Praat: doing phonetics by computer*. Computer program.

[2] McCauley, S. M., Hestvik, A., & Vogel, I. 2012. Perception and bias in the processing of compound versus phrasal stress: Evidence from event-related brain potentials. *Language and Speech*, 56(1), 23-44.

[3] Morrill, T. 2011. Acoustic correlates of stress in English adjective-noun compounds. *Language and Speech*, 55(2), 167-201.

[4] Vogel, I., & Raimy, E. 2002. The acquisition of compound vs. phrasal stress: The role of prosodic constituents. *Journal of Child Language*, 29(2), 225-250.

Perceiving and producing clusters in French: the case of Japanese learners

Rachel Albar, Hiyon Yoo
UMR 7110/ Laboratoire de linguistique formelle, Université Paris-Diderot, France

In Japanese, the phonological structure does not allow initial clusters nor consonants in coda position (except in some cases the nasal /N/, [1]) while in French, we can find up to three consonants in onset position. Therefore, in order to repair illicit clusters in the phonological structure of their mother tongue, Japanese learners of French insert an epenthetic vowel between consonants and after a consonant in coda position. In a previous study, Dupoux et al. [2,3] have shown that Japanese speakers perceive more epenthetic vowels than French, in non-word synthetic stimuli but also in naturally produced non-words where coarticulation is kept. Detey [4] arrives at the same conclusions and goes further, showing that some combinations of consonants such as an obstruent followed by a nasal or plosive, induce more epenthesis.

The aim of this presentation is to see how Japanese learners of French deal with clusters, in perception but also in production. Our hypotheses concern the learner's proficiency level, the nature of the cluster but also the nature of the epenthetic vowel, and for the production part, the type of production (reading and semi-spontaneous speech). We predict that:

1. There is a correlation between the production and the perception of epenthetic vowels and the level of performance, advanced learners having better performance than beginners
2. There is a correlation between the nature of consonants and epenthesis; Following Tajima et al. [5] voicing should induce more epenthesis, and more epenthesis should occur with consonant associated with /l/ in comparison to /r/ because of its proximity to the liquid in Japanese /ɾ/.
3. In production, the inserted and perceived epenthesis would be the same inserted in loanwords, usually [ɯ], and [o] after [t] or [d].
4. For production, we predict that the semi-spontaneous task induces more epenthesis, especially in advanced learners.

The perception experiment consisted of a discrimination test following the AXB paradigm with triplets of non-words. The stimuli were non-words following the pattern C1(V)C2a, with different plosives for C1, and /l/ or /r/ for C2. The vowels were [a], [u] and [ø]. We obtained pairs such as bla/bala. Native speakers of French produced the stimuli in carrier sentences.

For production, we wanted to make a comparison between productions in Japanese and in French. We thus built a corpus that consisted of loanwords with word initial consonant clusters and city names. The target words were put in sentences, a short text but also a map task, in order to have different type of productions (read and semi-spontaneous productions)

Results show that there is a great correlation between the nature of consonants and epenthesis. Epenthesis occurs in production with voiced consonant especially for clusters combined with /r/, while unvoiced clusters induced less epenthesis in production and perception. Our hypothesis concerning the difficulty of /l/ due to its assimilation to the Japanese liquid was confirmed, especially in production where more epenthesis was produced with /l/ than /r/. The same tendency could be noted in perception, but only for some clusters. Acoustic analyses of inserted vowels confirmed the third hypothesis: epenthesis vowel was generally the vowel [ɯ] in production and we observed a more important error rate in perception when the stimuli used [ø] epenthesis, acoustically close to the Japanese [ɯ].

However, results show no significant correlation between epenthesis and level of proficiency even if advanced learners have a tendency to perceive and produce less epenthesis. The type of task (reading/map task) was also not significant.

[1] Kubozono H., 2015, Introduction to Japanese phonetics and phonology. In Kubozono, H. (Ed.). 2015. *Handbook of Japanese Phonetics and Phonology vol. 2*. Walter de Gruyter GmbH & Co KG. 1-42.

[2] Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., & Mehler, J. 1999. Epenthetic vowels in Japanese: A perceptual illusion? *Journal of experimental psychology: human perception and performance*, 25(6).

[3] Dupoux, E., Parlato, E., Frota, S., Hirose, Y., & Peperkamp, S. 2011. Where do illusory vowels come from? *Journal of Memory and Language*, 64(3), 199-210.

[4] Detey, S., 2005. *Interphonologie et représentations orthographiques. Du rôle de l'écrit dans l'enseignement/apprentissage du français oral chez des étudiants japonais* (Doctoral dissertation, Université Toulouse le Mirail-Toulouse II).

[5] Tajima, K., Erickson, D., & Nagao, K. 2014. Production of syllable structure in a second language: Factors affecting vowel epenthesis in Japanese-accented English. *IULC Working Papers*, 2(2).

Conflicting functions and phonetic cues in bilingual Spanish and Quechua

Timo Buchholz & Uli Reich
Freie Universität Berlin

Culminativity, the time-honored cornerstone of Metrical Phonology [1, 2], appears quite porous if looked at from a wider comparative perspective. While many European languages clearly have one and only one most prominent syllable in the phonological word or phrase that also accumulates the prosodic cues time, tone, formant structure and intensity, tonal languages distribute tones independently from a stressed syllable [3], languages like Japanese or Wolof show free distinctive length [4, 5], and Portuguese shows stress clearer than Spanish [6]. In view of similar facts from a wide range of languages, [7] advocates for a property-driven prosodic typology that refrains from the search for "pigeon-hole"-types. In our view, prosodic diversity is due to the conflict of four antagonist functions that time, f0, formant structure and intensity can serve: *distinctivity* prefers the free distribution of time and tone and thus diminishes the salience of a special syllable *culminativity* assigns stress to. The towering dominance of non-peripheral stress also suffers from prosodic and segmental processes that serve *delimitation.* The fourth function, absent in Trubetzkoy's reasoning, is *rhythmicity*, a musical principle that assigns alternating prominence at different prosodic domains [6]. Languages and utterances differ in the particular constellation these four conflicting functions are balanced in. Our contribution tries to show this trade off at work in the prosodic form of utterances in a bilingual speech community.

The data we present here consists of a corpus of semi-spontaneous dialogical speech elicited via maptasks and re-tellings of stories with metrically controlled content words in Spanish and Quechua, gathered during a fieldwork trip in the summer of 2015 in Conchucos, Peru. Based on a quantitative phonetic study, we observe different acoustic reduction (centralization, devoicing [8, 9, 10] and complete elision of vowels, assimilation and elision of consonants) and lengthening phenomena (of vowels) in the Spanish of ten speakers. We propose to analyze them as reflecting the conflicting demands of lexical prominence, rhythmicity and prosodic phrasing, with a high degree of correspondence between the different actual phonetic means and these phonological functions. Since timing differences are distinctive in the local variety of Quechua, we propose to use moras also in our analysis of the Spanish data. In a second step, we relate the variation in these prosodic timing phenomena [11] further to the sociolinguistic background of the speakers, especially their relative dominance of Quechua and Spanish. We propose that the local variety of Quechua does not have lexical stress, just as is argued for e.g. French and Korean by [12] and [13], respectively, and that tonal movement is only employed to mark boundaries at the level of the phrase. Lengthening and reduction are employed in our Spanish data variably to emulate two distinct ideal rhythmic types, oscillating between a more Spanish-dominant speaking style and one that is more Quechua-dominant, broadly speaking. Spanish-dominant speaking style aims at producing rhythmic sequences of units (syllables) of equal length and alternating prominence ("pasas por debajo del perro" *you pass below the dog* is realized as [ˈpa:z.pɔ.rə.ˈβa:.xəð.ˈpɛ:.rẹ], with a structure of ˈμμ.μ.μ.ˈμμ.μμ.ˈμμ.μ), whereas Quechua-dominant speaking style aims for sequences of complex syllables of equal length and prominence ("con su mano le tiene de ahí dice" *in his hand there he has, they say* is realized as [ˌkõs.ˈma:n.lə.ˈt͡jen.ˈd͡j͡aj:.ˈdi:s], with a structure of ˌμμ.ˈμμ.μ.ˈμμ.ˈμμ.ˈμμ). While it is only the presence / absence of the property of relative prominence that distinguishes these ideal types, we emphasize that 1) both are the result of a complex interplay of prosodic constraints whose origin cannot easily be attributed to either "Spanish" or "Quechua" in this particular multilingual setting and 2) their use is subject to communicative goals and social stances with local meaning [14].

[1] Trubetzkoy, Nikolai S. 1939. Grundzüge der Phonologie: Publié avec l'appui du Cercle linguistique de Copenhague et du ministère de l'instruction publique de la République tchéco-slovaque. Prague.

[2] Hayes, Bruce. 1995. Metrical stress theory: Principles and case studies. Chicago: University of Chicago Press.

[3] Yip, Moira J. W. 2002. Tone. Cambridge/ New York: Cambridge University Press.

[4] Beckman, Mary E. 1986. Stress and non-stress accent. Dordrecht: Foris.

[5] Ka, Omar. 1994. Wolof Phonology and Morphology. Lanham: University Press of America.

[6] Reich, Uli & Martin Rohrmeier. 2014. Batidas latinas: On rhythm and meter in Spanish and Portuguese and other forms of music. In Javier Caro Reina & Renata Szczepaniak (Eds.), *Syllable and word languages*. Berlin: de Gruyter, 391–420.

[7] Hyman, Larry M. 2014. Do All Languages Have Word Accent? In Harry van der Hulst (Ed.), *Word stress: Theoretical and typological issues*. Cambridge: Cambridge University Press, 56–82.

[8] Beckman, Mary E. 1996. When is a syllable not a syllable? In Takashi Otake & Anne Cutler (Eds.), *Phonological structure and language processing: Cross-linguistic studies*. Berlin, New York: Mouton de Gruyter , 95–123.

[9] Delforge, Ann M. 2008. Unstressed Vowel Reduction in Andean Spanish. In Laura Colantoni & Jeffrey Steele (Eds.), *Selected Proceedings of the 3rd Conference on Laboratory Approaches to Spanish Phonology*. Somerville, MA: Cascadilla Proceedings Project, 107–124.

[10] Delforge, Ann M. 2011. Vowel Devoicing in Cusco Collao Quechua. *Proceedings of the 17th International Congress of Phonetic Sciences*, (Hong Kong), 556–559.

[11] Fletcher, Janet. 2010. The Prosody of Speech: Timing and Rhythm. In William J. Hardcastle, John Laver & Fiona E. Gibbon (Eds.), *The Handbook of phonetic sciences, 2nd edn*. Oxford: Blackwell Publishers, 523–602.

[12] Jun, Sun-Ah & Cécile Fougeron. 2000. A Phonological Model of French Intonation. In Antonis Botinis (Ed.), *Intonation: Analysis, Modelling and Technology*. Dordrecht: Springer, 209–242.

[13] Jun, Sun-Ah. 2005. Korean Intonational Phonology and Prosodic Transcription. In Sun-Ah Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing*. Oxford, New York: Oxford University Press, 201–229.

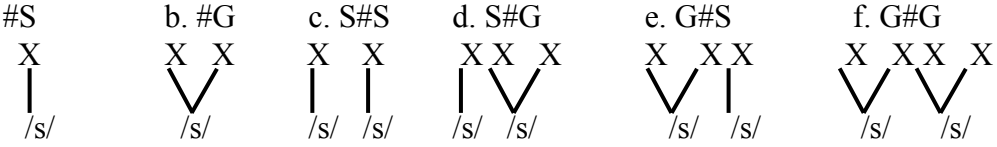
[14] Mufwene, Salikoko S. 2001. The ecology of language evolution. Cambridge / New York: Cambridge University Press.

On the phonetics of four-level length contrast: Evidence from Moroccan Arabic

Giuseppina Turco[1], Karim Shoul[1], Rachid Ridouane[1]
[1] *Laboratoire de Phonétique et Phonologie (UMR 7018) – CNRS/Paris 3*

This study investigates the temporal and non-temporal characteristics of identical consonant sequences in Moroccan Arabic. The phonological contrasts examined involve sequences across word boundaries, opposing singleton (S) and geminate (G) dental stops and fricatives in six contexts: #S, #G, S#S, S#G, G#S, and G#G (see Table 1). The underlying representations of these sequences are shown in (1):

(1)



The comparison of the six sequences seeks to determine the way these representations are reflected in the acoustic details of speech production. Specifically, we investigate whether the four-level length contrast displayed in (1) (i.e. from 1 to 4 timing slots) is maintained at a phonetic level, and how far the nature of the sequence as well as the manner of articulation of the consonants shape variability in consonant duration. The duration of the relevant consonants was measured in Praat. Non-temporal boundary cues were also investigated, and the presence or absence of releases, transitional schwas and pauses was noted. The difference between different types of sequences was evaluated by using mixed effects models.

Results from non-temporal parameters show that when two identical fricatives are adjacent, they are produced as one long uninterrupted frication noise, regardless of the singleton/geminate nature of the combined consonants. When two stops are adjacent, the transition between them may take two forms. Their oral closures may merge into one long closure phase, or the closure of the first stop may be released before that of the following stop is formed. As shown in Table 2, the tendency to release the first stop varies depending on whether or not the second stop is a geminate. In S#G and G#G sequences, speakers overwhelmingly chose to release the first stop. When the second stop is a singleton (S#S and G#S), speakers chose much more often to produce the two stops as a single long closure phase, similar to a lexical geminate #G. Results from the durational data (see figure 1 for fricatives) show that S#S has virtually the same duration as a lexical geminate #G (β=-4.9, SE=4.9, t=-1, p=.3), supporting their identical representation as two timing slots at the prosodic level. When a geminate consonant is included within the sequence, two patterns emerge, depending again on whether or not the second consonant is a geminate. Mirroring the non-durational observations, S#G and G#G pattern together and exhibit significantly longer durations compared to G#S ($\beta_{S\#G}$=59, SE=5.2, t=11.2, p<.0001; $\beta_{G\#S}$=69.9, SE=5.8, t=12, p<.0001), S#S ($\beta_{S\#G}$=76.4, SE=4.8, t=15.8, p<.0001; $\beta_{G\#S}$=87.3, SE=5.4, t=15.9, p<.0001) and #G ($\beta_{S\#G}$=71.4, SE=4.9, t=14.4, p<.0001; $\beta_{G\#S}$=82.3, SE=5.5, t=14.8, p<.0001). The fact that G#G does not display a much longer duration than S#G (11 ms on average, β=10.9, SE=5.4, t=1.9, p=.04) and that G#S is only slightly longer that S#S (17 ms, β=17.3, SE=5.2, t=3.3, p=.001), and #G (12 ms, β=12.3, SE=5.3, t= 2.3, p=.05) points towards a temporal reorganization, reflected mainly in the much shorter duration of the first consonant when the second consonant is a geminate (figure 2).

Taken together these results suggest that the durational hierarchy within categories is limited by a constraint on the number of levels of length. Three degrees of length, even across word boundaries, are possibly the upper limit at the production level (i.e. S<G=SS=GS<SG=GG). Because it involves too much crowding in the consonant duration space, it is expected that no language may have a four-level length contrast. And, as we shall argue, three-level length systems may need to supplement duration by some non-durational features to allow listeners to reliably perceive the contrast between the different consonant lengths.

Table 1. Speech data analyzed. Each phrase was embedded within the carrier sentence *gal … ʒuʒ mrrat* "He said … twice" and produced by three native speakers of Moroccan Arabic. Note that when one of the consonants in the sequence is emphatic /Cˤ/ the entire sequence may be emphatic.

	#S	#G	S#S	G#S	S#G	G#G
stops	ha dat	ha ddat	ʕad dab	ʕadd dat	had ddat	ʕadd ddat
	ha dˤar	ha dˤdˤar	ʕad dˤar	ʃadd dˤar	had dˤdˤar	ʃadd dˤdˤar
fricatives	ha saf	ha ssaf	ras saf	ʕass saf	ras ssaf	ʕass ssaf
	ha sˤak	ha sˤsˤak	qas sˤak	qas sˤsˤak	xasˤsˤ sˤak	xasˤsˤ sˤsˤak

Table 2. Number of sequences produced with or without release for dental stop sequences.

	#S	#G	S#S	G#S	S#G	G#G
Unreleased	34	32	33	28	6	3
Released	0	0	1	17	34	47
Pause			*0*	*4*	*0*	*9*
Schwa			*0*	*12*	*24*	*32*
Total	**34**	**32**	**34**	**45**	**40**	**50**

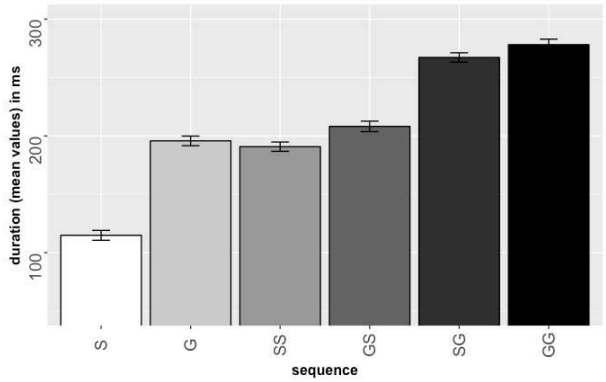


Figure 1. *Mean duration of the dental fricatives as function of combination type (#S, #G S#S, G#S, S#G, G#G). Whiskers represent standard errors as calculated from the model.*

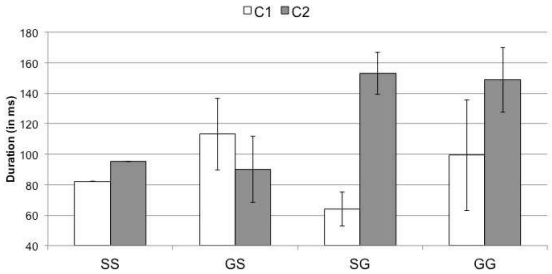


Figure 2. *Mean (raw) duration and standard deviations of the first (C1) and second (C2) dental stops as function of sequence type (S#S, G#S, S#G, G#G). For the sequence S#S only one occurrence with release was produced (see Table 2 for the other sequences).*

Perceptual voicing asymmetries in Dutch and English native listeners

Antje Stoehr[1,2], Jessica Zurlo[3] & Janet G. van Hell[3]
[1]*Radboud University,* [2]*IMPRS for Language Sciences,* [3]*The Pennsylvania State University*

Infants tune in to their native language during their first year of life: They lose the ability to distinguish non-native speech sounds, and maintain the ability to discriminate native speech sounds [1]. However, the discrimination of certain native speech sounds is asymmetric: For example, Dutch and Japanese infants recognize that a /p/ may not be mispronounced as a [t], but they do not recognize a mispronunciation of /t/ as [p] [2]. Similar asymmetries last into adulthood [3].

According to the Featurally Underspecified Lexicon model [4], acceptability of [p] as a good candidate of /t/ results from underspecification of the coronal place of articulation in /t/. While asymmetries in the perception of place of articulation are well established, asymmetries in the perception of voicing remain controversial.

The voicing contrast between voiced stops /bdg/ and voiceless stops /ptk/ is present in most of the world's languages [5]. However, the concrete phonetic implementation of voicing differs cross-linguistically [6]. In Dutch, voiced stops are prevoiced, and voiceless stops have short lag VOT. In English, voiced stops have short lag VOT and voiceless stops are aspirated.

Some researchers consider voiceless stops underspecified [4; 7], but others consider short lag stops, which can be voiced or voiceless depending on the language, underspecified [8; 9]. These two views make different predictions for voicing asymmetries in prevoicing languages and aspiration languages. If voiceless stops are underspecified in all languages, English and Dutch speakers should likewise accept a mispronunciation in the direction voiceless→voiced (/p/→[b]), but not vice versa. If instead short lag is underspecified, English listeners should accept a mispronunciation in the direction short lag→aspiration (/b/→[p]), but not vice versa; and Dutch listeners should accept a mispronunciation in the direction short lag→prevoicing (/p/→[b]), but not vice versa. Lastly, previous psychological research identified perception-driven mechanisms [10], which predict that omissions of acoustic signals are undetected and thus acceptable, suggesting that mispronunciations in the direction prevoicing→short lag (Dutch: /b/→[p]) and aspiration→short lag (English: /p/→[b]) are acceptable. By contrast, addition of acoustic signals should represent an inacceptable mispronunciation, suggesting that mispronunciations in the direction short lag→prevoicing/aspiration should be detected.

To test these three hypotheses, 24 Dutch listeners and 13 American English (AE) listeners (data collection ongoing, goal N=24) completed a mispronunciation detection task. Participants saw pictures and heard words in which word-initial stops were either correctly pronounced or mispronounced (/b/→[p] and /p/→[b]). The stimuli were naturally produced words adjusted for VOT duration. Mispronunciations did not result in existing words.

Dutch and AE listeners differed in their acceptability judgment of the mispronunciations (Figure 1): Dutch listeners judged a mispronunciation in the direction prevoiced→short lag (/b/→[p]) acceptable, but not a mispronunciation in the direction short lag→prevoiced (/p/→[b]) ($\beta_{Condition}$=-1.75, SE=0.13, z=-13.38, p<.001; $\beta_{Voicing}$=-1.35, SE=0.13, z=-10.70, p<.001; $\beta_{Condition*Voicing}$=-1.12, SE=0.07, z=-15.76, p<.001). AE listeners, by contrast, judged a mispronunciation in the direction voiceless→voiced (/p/→[b]) more acceptable than a mispronunciation in the direction voiced→voiceless (/b/→[p]) ($\beta_{Condition}$=-2.87, SE=0.45, z=-6.39, p<.001; $\beta_{Voicing}$=0.61, SE=0.19, z=3.20, p=.001; $\beta_{Condition*Voicing}$=0.54, SE=0.14, z=3.78, p<.001).

These results support perception-driven mechanisms, in which acoustic information may be omitted, but not added [10]. Specifically, Dutch and AE listeners accept mis-pronunciations resulting in short lag VOT as acceptable realization for target prevoiced (Dutch) and target aspirated (AE) stops, but reliably detect the reverse mispronunciations.

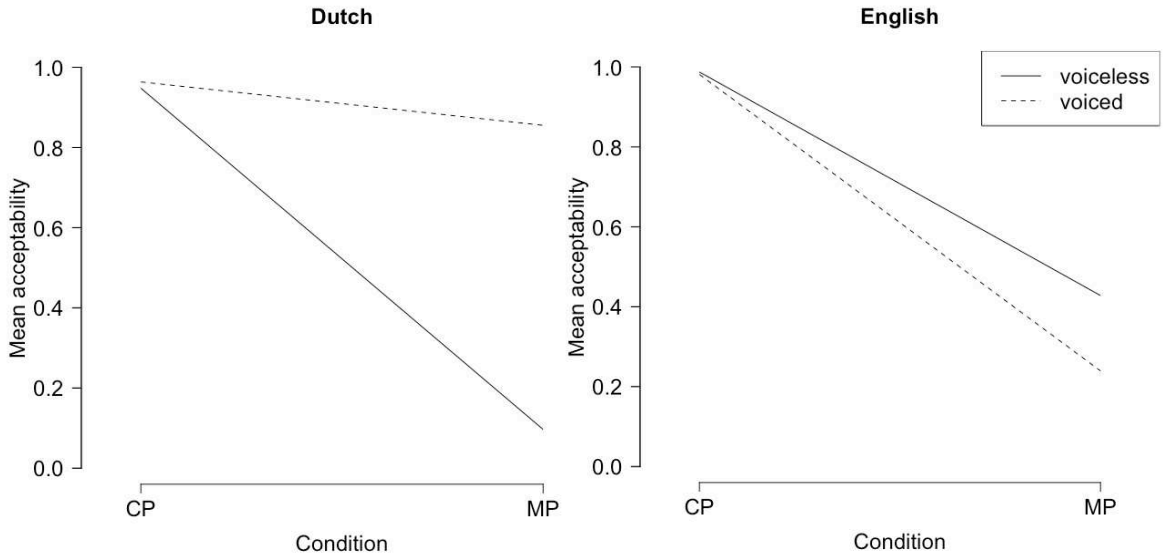


Figure 1. *Dutch (left) and English (right) native speakers' acceptability judgments for correct pronunciations (CP) and mispronunciations (MP) by voicing categories*.

[1] Kuhl, P.K. 2006. Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, 5, 831-843.

[2] Tsuji, S., Mazuka, R., Cristia, A., & Fikkert, P. 2015. Even at 4 months, a labial is a good enough coronal, but not voice versa. *Cognition*, 134, 252-256.

[3] Cornell, S.A., Lahiri, A., & Eulitz, C. 2013. Inequality across consonantal contrasts in speech perception: Evidence from mismatch negativity. *Journal of Experimental Psychology: Human Perception and Performance*, 39, 757-772.

[4] Lahiri, A., & Reetz, H. 2010. Distinctive features: Phonological underspecification in representation and processing. *Journal of Phonetics*, 38, 44-59.

[5] Ladefoged, P., & Maddieson, I. 1996. *The sounds of the world's languages*. Oxford: Blackwell.

[6] Lisker, L., & Abramson, A. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 20, 384-422.

[7] Hwang, S.-O.K., Monahan, P.J., & Idsardi, W.J. 2010. Underspecification and asymmetries in voicing perception. *Phonology*, 27, 205-224.

[8] Hestvik, A., & Durvasula, K. 2016. Neurobiological evidence for voicing underspecification in English. *Brain & Language*, 152, 28-43.

[9] Kager, R., van der Feest, S., Fikkert, P., Kerkhoff, A., & Zamuner, T. 2007. Representations of [voice]: Evidence from acquisition. In van de Weijer, J., & van der Torre, E. (Eds.), *Voicing in Dutch: (De)voicing phonology, phonetics, and psycholinguistics.* Amsterdam: Benjamins, 41–80.

[10] Gilmore, G.C., Hersh, H., Caramazza, A., & Griffin, J. 1979. On the prediction of confusion matrices from similarity judgments. *Perception & Psychophysics*, 25, 425-431.

Dissimilation can be gradient: evidence from Aberystwyth English

Adèle Jatteau[1] and Míša Hejná[2]
[1]*Université Paris 8*, [2]*Aarhus University*

Dissimilation is classically considered a phonetically abrupt process, both as a synchronic rule (e.g. Suzuki 1998) and as a sound change (Hock 1991): a feature is categorically removed in the vicinity of the same feature in a given domain. Contrary to these expectations, this study presents evidence for a phonetically gradient pattern of aspiration dissimilation found in Aberystwyth English (Wales): an aspiration feature is *reduced* in the vicinity of the same feature. The process is comparable to two other patterns of gradient dissimilation recently reported in the literature, Mongolian (Svantesson et al. 2005, Svantesson & Karlsson 2012) and Georgian (Begus 2016). Together, these patterns show that phonetic investigation allows for a reanalysis of well-known phonological processes, and raise in turn the question of their phonetic motivation.

According to Ohala (e.g. 1981) and Blevins & Garrett (1998), aspiration is a feature with 'stretched-out cues': it is realized over a long temporal interval. Fortis stops in Aberystwyth English (AE) are both pre- and post-aspirated, as in *weapon* ['wɛ̝^{h}p^{h}ə̝n] (Hejná 2015) (see also Fig. 1 here); the same has been found for Manchester English (Hejná & Scanlon 2015) and Welsh (Morris & Hejná, in prep; Iosad, pc). CVC structures in which the intervening vowel is surrounded by fortis stops, as in *tip* [t^{sh}ɪ^{h}p^{h}], or by /h/ and a fortis stop, as in *hip* [hɪ^{h}p^{h}], are then most likely to present coarticulation of aspiration throughout the vowel. In contrast with these predictions, the aspiration feature of stops in Aberystwyth English (AE) is gradiently reduced when preceded by another aspirated stop or /h/ ('SG' in Figures). Phonetic measurements based on 12 English speakers from Aberystwyth (550-620 tokens per speaker) show that the degree of pre-aspiration of C_2 in C_1VC_2- sequences depends on the quality of C_1: the intervening vowel shows less breathiness and pre-aspiration, and shorter intervals of breathiness and pre-aspiration, when C_1 is a *fortis* stop (/p, t, k/, as in *tip*) or /h/ than when it is a *lenis* stop (/b, d, g/, as in *dip*) or a sonorant (as in *lip*). The difference is significant for both breathiness and pre-aspiration (Fig. 2 and 3), and consistent across words and prosodic contexts. Measurements of the breathiness of the intervening vowel also show that complete coarticulation between both aspiration features is very rare. Analyses of the overall noisiness of the vocalic interval will be conducted via Cepstral Peak Prominence in order to evaluate the degree of coarticulation via noisiness level (rather than just durational and occurrence-related properties). Preliminary analyses finally suggest that the pattern is sensitive to vowel length. There is thus a pattern of progressive, phonetically gradient, and lexically regular dissimilation of aspiration in AE.

The gradient reduction of aspiration does not seem to be limited to AE: two other comparable patterns are reported in the literature in unrelated languages. In Halh Mongolian, the post-aspiration of C_1 in C_1VC_2- sequences is significantly shorter before pre-aspirated stops than before other consonants: $\underline{C}^{h}V^{h}C$- (Svantesson et al. 2005, Svantesson & Karlsson 2012). In Georgian, where aspirated stops are post-aspirated, Begus (2016) finds a reduction of the aspiration feature of C_2: $C^{h}V\underline{C}^{h}$. While it is unclear now whether these three patterns stem from the same cause, they suggest that gradient dissimilations of aspiration features are well attested cross-linguistically.

These patterns have interesting implications for diachrony: several arguments suggest that gradient dissimilation could be a(nother) phonetic precursor to the completed dissimilations reconstructed for example in Sanskrit or other dialects of Mongolian. The question is then what motivates the gradient reduction. There are broadly speaking two accounts of dissimilation in the literature: Ohala's theory of dissimilation by hypercorrection (e.g. 1981), and the analysis of dissimilation as a type of speech error, which Garrett & Johnson (2012) and Garrett (2015) classify as motor planning or gestural organization problems. We discuss the question whether the AE data fits the predictions made by these two models and their variants.

Fig. 1. Distinction between breathiness (‘br’) and pre-aspiration (‘pre’).

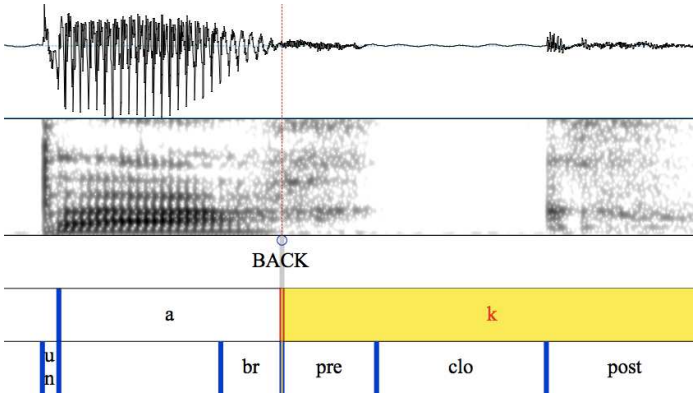


Fig. 2. Gradient dissimilation related to the occurrence of pre-aspiration and breathiness.

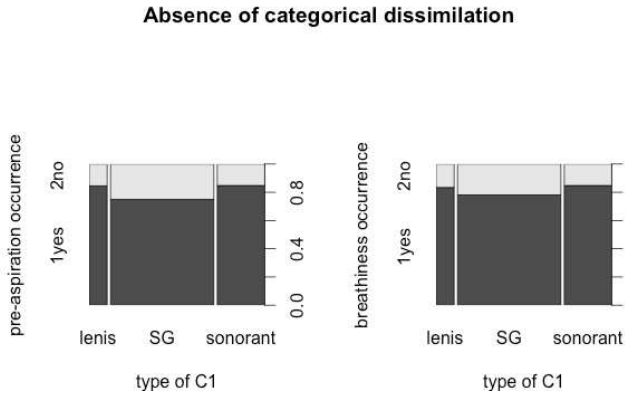
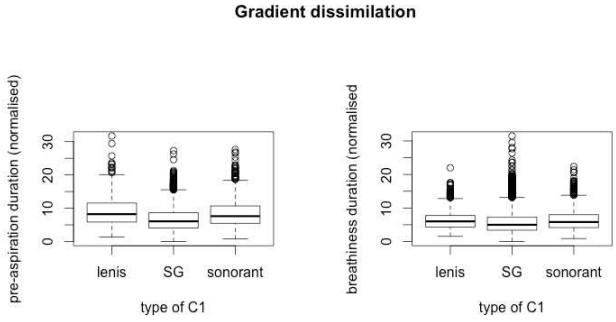


Fig. 3. Gradient dissimilation related to the duration of pre-aspiration and breathiness.



References

Begus, G. 2016. “The Phonetics of Aspirate Dissimilation: Grassmann’s Law in Georgian.” Poster. The South Caucasian Chalk Circle, Paris.

Blevins, J., and A. Garrett. 1998. “The Origins of Consonant-Vowel Metathesis.” *Language* 74 (3): 508–56.

Garrett, A. 2015. “Sound Change.” In *The Routledge Handbook of Historical Linguistics*, C. Bowern and B. Evans (eds). London: Routledge.

Garrett, A. and K. Johnson. 2012. “Phonetic Bias in Sound Change.” In *Origins of Sound Change: Approaches to Phonologization*, A. Yu (ed.), 51–97. Oxford: OUP.

Hejná, M. 2015. *Pre-Aspiration in Welsh English: A Case Study of Aberystwyth*. PhD thesis, University of Manchester.

Hejná, M. and J. Scanlon. 2015. “New Laryngeal Allophony in Manchester English.” *18th ICPhS*, Glasgow.

Hock, H. 1991. *Principles of Historical Linguistics*. 2nd ed. Berlin: de Gruyter.

Morris, J. and M. Hejná. In prep. “Pre-aspiration in Bethesda Welsh: a sociophonetic analysis”.

Ohala, J.J. 1981. “The Listener as a Source of Sound Change.” In *Papers from the Parasession on Language and Behavior*, C.S. Masek, R.A. Hendrick and M.F. Miller (eds), 178–203. Chicago: Chicago Linguistics Society.

Suzuki, K. 1998. “A Typological Investigation of Dissimilation.” PhD Dissertation, University of Arizona.

Svantesson, J.-O. and A. Mukhanova Karlsson. 2012. “Preaspiration in Modern and Old Mongolian.” *Suomalais-Ugrilaisen Seuran Toimituksia* 264: 453–64.

Svantesson, J.-O., A. Tsendina, A. Karlsson and V. Franzen. 2005. *The Phonology of Mongolian*. Oxford: Oxford University Press.

Wednesday, 13:15 - 14:15
Poster Session 5, #10

Imitation evidence for the encoding of prosodic detail in prenuclear accent patterns

Jennifer Cole[1], José I. Hualde[2], Suyeon Im[2]
[1]Northwestern University, [2]University of Illinois at Urbana-Champaign

In English, pitch accents mark focus and information status (IS) [1,2], yet semantic and pragmatic factors alone do not uniquely determine accent patterns. Even under explicit focus or IS conditions accent placement and melody are variable across speakers [3, 4; see 5 for German). Further, "ornamental" accents can optionally occur in prenuclear position on words that do not express focus or new information [6, 7]. We ask about the status of prenuclear pitch accents (PPA) in the cognitive representation of heard speech: If PPAs are variably present and largely ornamental, are they disregarded by the listener as prosodic "detail", or are they assigned a phonological representation and stored in the memory of a heard utterance? We probe listeners' reckoning of PPAs through the analysis of imitated speech. Under the hypothesis that listeners perceive and encode prosodic detail, we predict that PPAs will be reproduced in a task of intentional imitation. Conversely, the failure to imitate PPAs is consistent with the hypothesis that prosodic detail is not reliably or faithfully encoded, in at least some instances.

Corpus study: We examine PPA in English words with two possible landing sites for a pitch accent. These are words with a secondary stress that precedes the primary (2-1), where a PPA accent may occur on the syllable with primary stress (*èlevátion, òptimístic*), or on the earlier secondary stress (*élevàtion, óptimìstic*), which we term "early high". Early high is described in relation to rhythm and the preference for pitch accent at the left edge of the prosodic phrase [8, 9, 10]. However, these factors do not accurately predict all occurrences of early high, which we demonstrate with data from the Boston Radio News Corpus: among 288 words (48 words*6 announcers) with the 2-1 stress pattern, only 77 (27%) were produced with early high. Though all speakers produced early high, only 3 words were produced with early high by all speakers. The corpus data confirm that PPA in 2-1 words does not interact with focus, which is encoded through the nuclear accent, and is not strongly predicted by the IS of the prenuclear word, and hence is less informative than the nuclear accent as concerns semantic and pragmatic meaning.

Imitation experiment: 33 speakers of American English were asked to imitate an aural stimulus, with the instruction to repeat what they heard the way the model speaker said it. In stimulus sentences the first content word of the subject NP had a lexical 2-1 stress pattern (*The University of Kenya*), which was recorded by an informed model speaker with three accent patterns on the 2-1 word: Early high (H* on the secondary stress), Primary (accent on the primary stress), and Unaccented (no accent on the 2-1 word). The three versions of the subject NP were spliced onto the same sentence continuation, e.g., *"… won the tournament that nobody thought they could win"*. Participants were divided in three groups. Each group heard 12 utterances for each of the three accent patterns, but heard each sentence in only one accent pattern. Imitations were coded as Early high, Primary, or Unaccented based on the location of the f0 peak over the subject NP.

Imitation results: Early high was reliably imitated by all participants (91% recall accuracy), with lower recall accuracy for Primary (59%) and Unaccented (61%) stimuli (Fig. 1). Remarkably, Early high was also the most frequently produced "error" for Primary and Unaccented stimuli. A closer look at the early high imitations reveals that while participants faithfully reproduce an f0 peak in the region preceding the primary stress, many imitated productions (40%) are realized with a pitch accent pattern that is phonologically distinct from that of the stimulus (Fig. 2). These findings support the hypothesis that prosodic detail associated with PPA is encoded, though with variable mapping to phonological form.

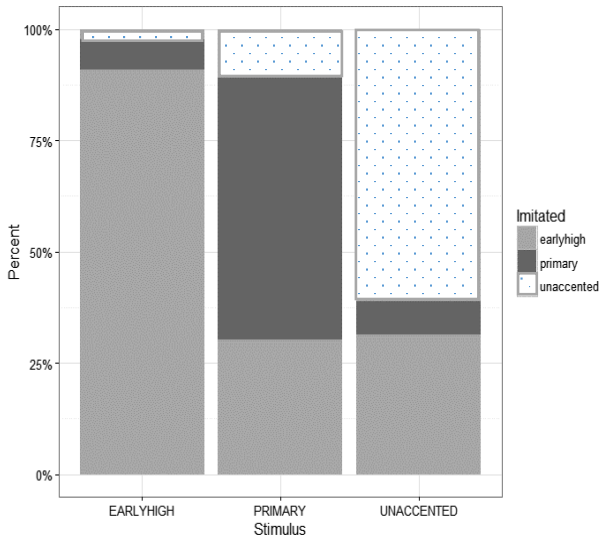


Figure 1. Number of imitated productions of stimuli grouped by prenuclear pitch accent (x-axis), with imitations coded as Early high (dotted white), Primary (dark gray) or Unaccented (light gray), based on location of f0 peak.

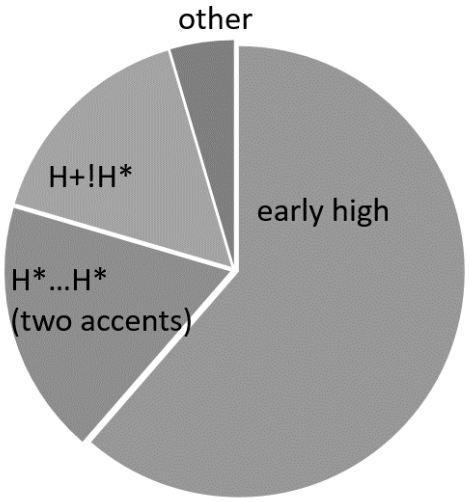


Figure 2. Early high imitated productions according to ToBI pitch accent label. Early high (H* on secondary); H+!H* on primary stress; H* on secondary and primary stresses; other patterns (data from 11 participants, ToBI labeling ongoing)

[1] Bolinger, D. 1958. A theory of pitch accent in English. Word, 14, 109–49.

[2] Ladd, D.R. 1980. The structure of intonational meaning: Evidence from English. Bloomington: Indiana University Press.

[3] Grabe, E. 2004. Intonational variation in urban dialects of English spoken in the British Isles. In Gilles, P. & Peters, J. (eds.), *Regional Variation in Intonation.* Tübingen: Niemeyer, 9–31.

[4] Yoon, T-J., 2010. Capturing inter-speaker invariance using statistical measures of rhythm. *Proceedings of the 5th Speech Prosody*, (Chicago).

[5] Mücke, D. & Grice, M., 2014. The effect of focus marking on supralaryngeal articulation–Is it mediated by accentuation? *Journal of Phonetics*, 44, 47–61.

[6] Calhoun, S. 2010. The centrality of metrical structure in signaling information structure: A probabilistic perspective. *Language*, 86, 1–42.

[7] Büring, D. 2007. Semantics, intonation and information structure. In Ramchand, G. & Reiss, C. (Eds.), *The Oxford Handbook of Linguistic Interfaces.* Oxford: OUP, 445–474.

[8] Shattuck-Hufnagel, S. 1995. The importance of phonological transcription in empirical approaches to "stress shift" versus "early accent": In Connell, B. & Arvaniti, A. (Eds.), *Phonology and Phonetic Evidence: Papers in Lab.Phonology IV.* Cambridge: CUP, 128–141.

[9] Grabe, E. & Warren, P., 1995. Stress shift: Do speakers do it or do listeners hear it. In Connell, B. & Arvaniti, A. (Eds.), *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV.* Cambridge: Cambridge University Press, 95–110.

[10] Gussenhoven, C. 2005. Procliticized phonological phrases in English: Evidence from rhythm. *Studia Linguistica*, 59(2–3), 174–193.

The Influence of Information Status on Prenuclear Accents in German

Stefan Baumann, Jane Mertens and Janina Kalbertodt
IfL Phonetik, University of Cologne, Germany

Background and Motivation: In the field of intonation research there are many studies dealing with the form, the function and the phonological status of *nuclear* accents, defined as the last fully-fledged pitch accent in an intonation unit. The position of the nuclear accent is generally claimed to determine the interpretation of an utterance, i.e. whether a broad or a narrow focus reading is intended or whether the last discourse referent in an utterance is marked as *given* (often indicated by deaccentuation) or *new* information (often accented). The status of *prenuclear* accents – i.e. pitch accents that occur *before* the nucleus within the same intonation unit – is less clear, however. It has often been claimed that prenuclear accents are irrelevant for the meaning of an utterance and that they may only be placed for rhythmic reasons (cf. Büring's [1] *ornamental accents*). Other studies, mostly investigating highly controlled lab speech, found that prenuclear accents were placed consistently, even on textually given information in contrastive contexts [2,3] or on (non-contrastive) topics in topic-comment structures [4]. However, the accents displayed subtle changes in peak scaling [2,4] or peak alignment [4], which expressed meaning-related differences. The aim of the present study is to learn more about the forms and functions of prenuclear accents in German by investigating the influence of a specific type of meaning, namely the information status of a sentence-initial referent, on its prosodic realisation. We expect to find a positive correlation between the level of newness/informativeness of a referent and its prosodic prominence.

Methods: Twenty native German speakers (13f, 7m), aged between 23 and 69, were presented with four different mini-stories on a computer screen. For each story, two pre-recorded context sentences were played to the subjects, who were asked to read out only the last sentence (printed in bold face) at a natural but swift speech rate (primed by a training item). By varying the second context sentence we designed four conditions rendering the subject argument in the target sentence either *given*, *accessible*, *new* or *contrastive* (see (1) for the target word *Superheld* 'superhero'; expected prenuclear and nuclear accents are underlined). Each participant was presented with only one information status condition, leading to five realisations per condition. The classification of phrase breaks and accent types that entered our analysis was based on a consensus judgment of three trained phoneticians.

Results: In 32.5 % of the cases (mostly if the target word had at least three syllables), subjects produced a phrase break. These were excluded because an accent on the sentence-initial target word becomes the nuclear accent of the intonation unit. All remaining 54 utterances carried a prenuclear accent on the target word, i.e. we found no cases of deaccentuation. The accent L*+H was the most frequent type in all conditions. Generally, however, the distribution of accent types indicates a slight increase in prominence from the encoding of *given* referents through *accessible* and *new* to *contrastive* items (see Fig.1; accent types are arranged according to their perceptual prominence in German [5]). This is mirrored, for instance, by an increasing percentage of rising pitch accents (L*+H and L+H*) in relation to low and high accents from *given* to *contrastive* information. Among the gradual phonetic cues which make up the accents, *pitch excursion* seems to be particularly important, although not for all target words alike. Figure 2 shows the results for *Banane* ('banana'), the only fully-voiced target word in the corpus: At least from *given* to *new* referents, all occurring accent types exhibit an increase in pitch excursion.

Conclusions: It could be shown that the information status of a sentence-initial referent affects the form of a prenuclear accent marking this referent. That is, *speakers seem to make systematic use of prenuclear accents to express meaning differences*. Our read data display the expected distribution for German (in tendency): The more informative a referent, the higher its prosodic prominence. Interestingly, however, we found that sentence-initial referents were always accented, even if they represented given information.

(1)
Context 1: Thomas ist gestern ins Kino gegangen. (*Tom went to the cinema last night.*)
Context 2a ('given'): Er hat sich einen Film über einen Superhelden angesehen. Die Geschichte war allerdings etwas ungewöhnlich. (*He watched a movie about a superhero. The story was a bit unusual, though.*)
Context 2b ('accessible'): Er hat sich einen Fantasy-Film angesehen. Die Geschichte war allerdings etwas ungewöhnlich. (*He watched a sci-fi fantasy movie. The story was a bit unusual, though.*)
Context 2c ('new'): Der Film hat ihm sehr gut gefallen. Die Geschichte war allerdings etwas ungewöhnlich. (*He liked the movie a lot. The story was a bit unusual, though.*)
Context 2d ('contrast'): Er hat sich einen Fantasy-Film angesehen, der etwas ungewöhnlich war. Der Bösewicht war edel. (*He watched a sci-fi fantasy movie that was a bit unusual. The villain was noble.*)
Target: Der **Su**perheld hat sich als Id**iot** erwiesen. (*The **su**perhero turned out to be an **id**iot.*)
[The other three target words were *Banane* 'banana', *Neffe* 'nephew' and *Chamäleon* 'chameleon']

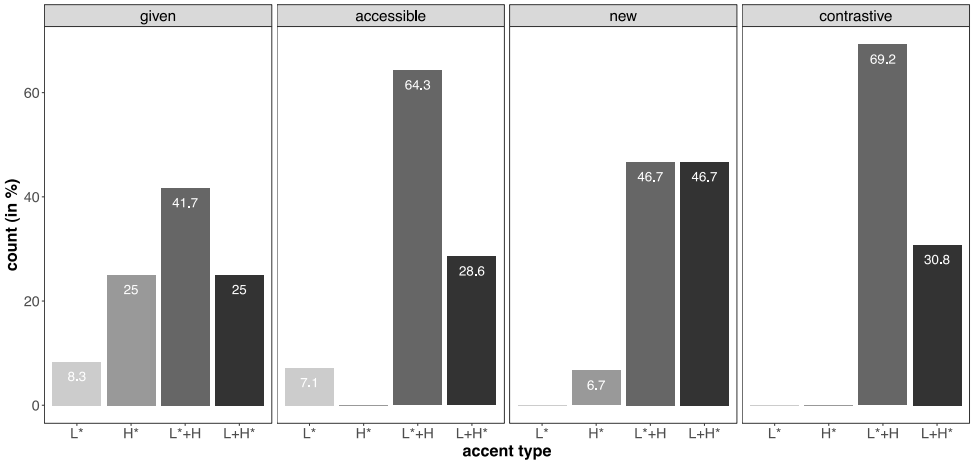


Figure 1. *Distribution of prenuclear accent types as a function of a sentence-initial referent's information status. Perceptual prominence of accent types increases from left to right.*

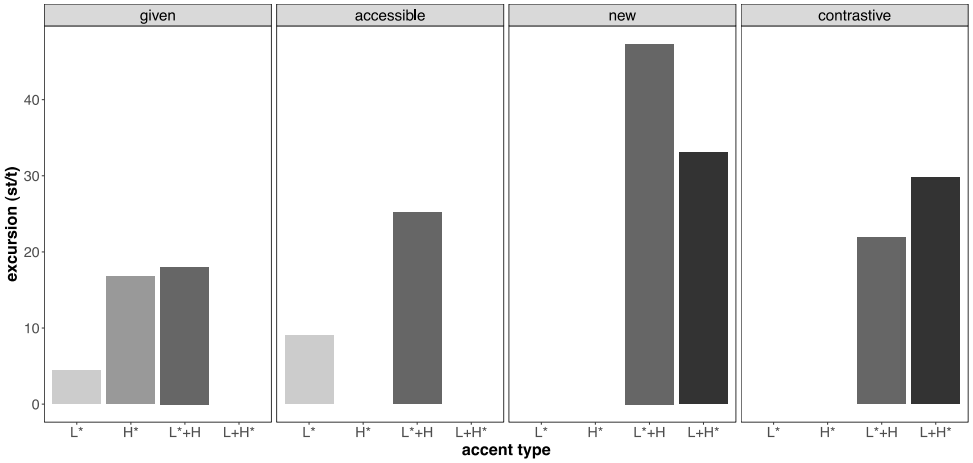


Figure 2. *Pitch excursion of prenuclear accents on the test word* Banane *('banana') as a function of its information status.*

[1] Büring, D. 2007. Intonation, Semantics and Information Structure. In Ramchand, G. & Reiss, C. (Eds.), *The Oxford Handbook of Linguistic Interfaces*. Oxford University Press, 445-474.
[2] Féry, C., & Kügler, F. 2008. Pitch accent scaling on given, new and focused constituents in German. *Journal of Phonetics* 36(4), 680-703.
[3] Baumann, S., Becker, J., Grice, M. & Mücke, D. 2007. Tonal and Articulatory Marking of Focus in German. *Proceedings of the 16th International Congress of Phonetic Sciences*, (Saarbrücken), 1029-1032.
[4] Braun, B. 2006. Phonetics and phonology of thematic contrast in German. *Language and Speech* 49(4), 451-493.
[5] Baumann, S., & Röhr, C. 2015. The perceptual prominence of pitch accent types in German. *Proceedings of the 18th International Congress of Phonetic Sciences*, (Glasgow), Paper number 298, 1-5.

Aerodynamic and articulatory explanations for the presence of nasal consonants in oral contexts in Bribri (Chibchan, Costa Rica)

Natacha Chevrier
Université Lyon 2, Laboratoire DDL

In Bribri – an endangered Chibchan language of Costa Rica – nasality is distinctive for vowels only. Nasal consonants ([m n ɳ̆ ɲ ɲ̆ w̃]) are the outcome of a leftward nasalization triggered by nasal vowels and targeting voiced consonants (/b ɾ d ɟ j w/) [1]. Nevertheless, some occurrences of nasal consonants cannot be explained by nasal spreading as no nasal vowels are present in the environment:

(i) Nasals in word final position (1)
(ii) Prenasalized consonants as the outcome of gemination (2)
(iii) Nasals which stand for voiceless stops assimilated in voicing (3)

Despite the existence of nasal consonants in oral contexts, it is still arguable that nasal consonants are not distinctive in Bribri. In this talk, I offer to explain the presence of nasal consonants in oral contexts based on aerodynamic and articulatory factors. The three cases listed above (i)-(iii) correspond to three processes:

(i) Hypervoicing of voiced stops in final position

It is well known that voiced stops are subject to the Aerodynamic Voicing Constraint [2]–[5]: the oral occlusion leads to an accumulation of the air in the oral cavity and to the quick equalization of pressures above and under the glottis, which in turn leads to the extinguishment of voicing. The AVC is higher in utterance initial and final positions, where the expiratory flow is low [5, pp. 153–157], [6, pp. 220–221]. There are two ways to adapt to the AVC [4]: let the stop get devoiced [7], [8] or use one of the several strategies of hypervoicing [3], [4], [9], [10]. In Bribri, the auditory salience of the onset (the "boost at onset" at the auditory nerve fibre level [11, pp. 43–45]) prevents stops from undergoing phonetic change in initial position. In final position however, hypervoicing is observed. There are several strategies of hypervoicing reported in the literature. Among them are nasalization [6], flapping [6, p. 224], [12] and retroflexion [13], [14]. Hypervoicing then allows to explain the presence of the nasal consonant in (1) as well as the [ɽ̆] realization of the coronal stop (/d/) in this same position (4).

(ii) Hypervoicing of long voiced stops

The AVC leads to the quick extinguishment of voicing. Hence, length and voicing of stops are hardly compatible. That explains why voiced geminated stops are typological rare and often undergo either devoicing or hypervoicing [15, pp. 320–324], [16, p. 89]. In Bribri, gemination resulting from morphological reduplication gives rise to long stops if voiceless (5) and to prenasalized consonants if voiced (2). A similar phenomenon is reported for Japanese [2, pp. 200–201]. Besides, length and nasality have been related in several studies [15, pp. 321–324], [17, pp. 23–24], [18]–[20].

(iii) Nasal release of 'unreleased' consonants

In Bribri, stops are preferably unreleased morpheme finally (6). In those cases, whereas the consonant is not orally released, it is actually sometimes released by the nasal cavity, leading to voiceless nasals (7). Voiceless nasals standing for unreleased consonants have previously been reported for other languages (e.g. Vietnamese) [21, p. 129].

In (3), the processes involved are as follow: 1. morpheme finally, the consonant is not released, leading to a voiceless nasal (t → n̥ / _+; k → ŋ̊ / _+); 2. the consonant of the following morpheme assimilates it in voicing (3)a (n̥ → n / _b) or in voicing and in place (3)b (ŋ̊ → m / _b).

While nasal harmony explains the majority of nasal consonants in Bribri [1], the remaining nasals can be explained by an aerodynamic (hypervoicing) and an articulatory (nasal released of 'unreleased' consonants) processes.

(1) a. [wîm] ‘man’[1]

b. [kɽûm] ‘plantain’

(2) a. [dí-n͡dí-ì] ‘very sharp’ (reduplication of [dí] with gemination of the consonant)

b. [bá-m͡bá-ì] ‘very hot’ (reduplication of [bá] with gemination of the consonant)

(3) a. [ˈbɛ́n-bɛ̄t-ì] ‘very light’ (reduplication of [bɛ́t])

b. [ˈám-bɽ-ùk] ‘to steal’ (composed of [ák] ‘theft’)

(4) a. [káɽ] ‘tree’

b. [îɽ] ‘brother, sister’

(5) a. [ʧɛ̃́-ˈt͡ʃɛ̃́-ì̀] ‘a lot’ (reduplication of [ʧɛ̃́] with gemination of the consonant)

b. [sɛ̃́-ˈsːɛ̃́] ‘cold’ (reduplication of [sɛ̃́] with gemination of the consonant)

(6) a. [ɟɨ́kⁿ] ‘nose’

b. [ˈdì-ű́kⁿ-tṵ̀] ‘dead person’

(7) a. [ʊ̄ˈpː-úpⁿ] ~ [ʊ̄ˈpː-úm̰] ‘light’

b. [bú-tⁿ] ~ [bú-n̰] ‘two’

References

[1] Constenla Umaña, 1985, “Prosodic nasality in Bribri and universals of nasality,” *Int. J. Am. Linguist.* 51(4), pp. 374–376; **[2]** Ohala, 1983, “The origin of sound patterns in vocal tract constraints,” in MacNeilage (ed) *The production of speech*, New York: Springer-Verlag, pp. 189–216; **[3]** Ohala, 1997, “Aerodynamics of phonology,” *Proceedings of the 4th Seoul International Conference on Linguistics (August 11-15)*, pp. 92–97; **[4]** Ohala, 2011, “Accommodation to the aerodynamic voicing constraint and its phonological relevance,” *Proceedings of the 17th International Congress of Phonetic Sciences (August 17-21)*, pp. 64–67; **[5]** Westbury & Keating, 1986, “On the naturalness of stop consonant voicing,” *J. Linguist.*, 22(1), pp. 145–166; **[6]** Solé, 2009, “Acoustic and aerodynamic factors in the interaction of features: The case of nasality and voicing,” in Vigário, Frota & Freitas (eds) *Phonetics and Phonology: Interactions and interrelations*, Amsterdam, Philadelphia: John Benjamins, pp. 205–234; **[7]** Myers, 2012, “Final devoicing: Production and perception studies,” in Borowsky, Kawahara, Shinya & Sugahara (eds) *Prosody Matters: Essays in Honor of Elisabeth Selkirk*, Sheffield, Bristol: Equinox, pp. 148–180; **[8]** Steriade, 2008, “The phonology of perceptibility effects,” in Hanson & Inkelas (eds) *The nature of the word*, Cambridge: MIT Press, pp. 151–179; **[9]** Westbury, 1983, “Enlargement of the supraglottal cavity and its relation to stop consonant voicing,” *J. Acoust. Soc. Am.*, 73(4), pp. 1322–1336; **[10]** Keyser & Stevens, 2001, “Enhancement revisited,” in Kenstowicz (ed) *Ken Hale: A life in languages*, Cambridge (Massachusett), London: MIT Press, pp. 271–291; **[11]** Wright, 2004, “A review of perceptual cues and cue robustness,” in Hayes, Kirchner & Steriade (eds.)*, Phonetically based phonology,* Cambridge, New York: Cambridge University Press, pp. 34–57; **[12]** Piñeros, 2003, “Accounting for the instability of Palenquero voiced stops,” *Lingua*, 113, pp. 1185–1222; **[13]** Sprouse, Solé & Ohala, 2008, “Oral cavity enlargement in retroflex stops,” *UC Berkeley Phonol. Lab Annu. Rep.*; **[14]** Hamann & Fuchs, 2008, “How do voiced retroflex stops evolve? Evidence from typology and an articulatory study,” *ZAS Pap. Linguist.*, 49, pp. 97–130; **[15]** Jaeger, 1978, “Speech aerodynamics and phonological universals,” *Proc. 4th Annu. Meet. Berkeley Linguist. Soc.*, pp. 311–329; **[16]** Ohala & Riordan, 1979, “Passive vocal tract enlargement during voiced stops,” *Speech Communication Papers, 97th Meeting of the Acoustical Society of America (June 12-16)*, pp. 89–92. ; **[17]** Sampson, 1999, *Nasal vowel evolution in Romance*. New York: Oxford University Press; **[18]** Hajek, 2003, “Patterns of vowel nasalisation in northern Italy: articulatory versus perceptual,” *Proceedings of the 15th International Congress of the Phonetic Sciences (August 3-9)*, pp. 235–238; **[19]** Whalen & Beddor, 1989, “Connections between nasality and vowel duration and height: elucidation of the eastern Algonquian intrusive nasal,” *Language (Baltim).*, 65(3), pp. 457–486; **[20]** Kohlberger, 2014, “Non-nasal nasals: Reinterpreting nasalisation in Shiwiar (Jivaroan),” *3rd Workshop on Sound Change in Interacting Human System (May 30)*; **[21]** Ladefoged & Maddieson, 1996, *The Sounds of the World’s Languages*. Oxford, Cambridge: Blackwell.

[1] The data presented in the examples were collected between 2012 and 2014 in the bribri villages of Bajo Coen and Amubre (Talamanca, Costa Rica).

Post-tonic consonant lengthening in Djambarrpuyŋu

Kathleen Jepson[1,2], Janet Fletcher[1,2] and Hywel Stoakes[1,2]
[1]*The University of Melbourne,* [2]*ARC Centre of Excellence for the Dynamics of Language*

Traditionally, Australian languages in the Pama-Nyungan language family are analysed as being stress languages where stress is fixed on the initial syllable of words, with alternating, trochaic meter, which is sensitive to morphological structure but not quantity (see [1]). Pitch has been found to be the major cue to main stress prominence in many Australian languages (see [2]), yet other acoustic correlates of prominence typically associated with vowels in many stress languages such as duration, peripheral vowel quality, and greater intensity are not found to be reliable cues to main stress. It has been claimed that consonants play a key role in signalling prosodic prominence in Australian languages (e.g. [3]). A consonant *following* a stressed vowel, regardless of whether it is in the stressed syllable or not, is usually strengthened or lengthened (see e.g. [4] for Warlpiri, [5] for Mawng). This phenomenon is often referred to as "post-tonic strengthening" [3] and is typically only recorded for the highest level of metrical prominence. It remains to be seen how widespread this pattern is across Indigenous Australian languages.

This study examines consonant duration as a cue to signalling main stress in Djambarrpuyŋu, an under-described Pama-Nyungan language spoken in northeast Arnhem Land, northern Australia. This study draws on language data consisting of words that vary in terms of segmental content, syllable length, and morphological complexity, produced in phrase initial, medial, and final position within a sentence frame, to investigate whether post-tonic lengthening is employed in Djambarrpuyŋu as a cue to metrical strength.

Recordings of four Djambarrpuyŋu speakers were made in Milingimbi, northern Australia. Sound files were segmented using MAUS [6] using a custom parameter model based on the SAMPA model, and manually corrected in Praat. Metrical strength was marked in the syllable strength tier according to phonological rules proposed for Djambarrpuyŋu [7]. A database was created in the EMU-webApp [8] and queried in R. The duration of stops and nasals are examined in word initial (i.e. tonic syllable onset), intervocalic post-tonic, and intervocalic non-post-tonic (elsewhere) position (n=2,157). The comparison between the post-tonic and the elsewhere condition is of primary interest. The linear mixed effects model included a fixed factor of the interaction between "phonemic length of preceding vowel" and "speaker", for a main effect of "consonant identity and position".

Results (see Figures 1,3) show that the velar stops /k, ɡ/, the lenis bilabial stop /b/, and the fortis retroflex stop /ʈ/ are significantly longer in post-tonic position than in other contexts, although other consonants e.g. the fortis palatal stop /c/ are not significantly affected by position within the word. Intervocalic nasal consonant duration is also not affected by word position in this data except for the velar nasal /ŋ/ (see Figures 2,3). Word initial consonants for the most part do not show position related lengthening despite being in the tonic syllable. The stop contrast is neutralised in word-initial position, whereas the full range of consonant contrasts are realised in post-tonic (and other medial) position; this may explain the patterns observed in Figure 3.

These results suggest lengthening of /b, k, ɡ, ʈ, ŋ/ in post-tonic position occurs in Djambarrpuyŋu. This may contribute to the perceptual prominence of the tonic syllable, and also be a way to maximise perceptual cues to place of articulation distinctions among the consonants of this Australian language (see [3]). Further results suggest that post-tonic stops and nasals which are in the C1 position of a cluster are also lengthened and this will be investigated in future work.

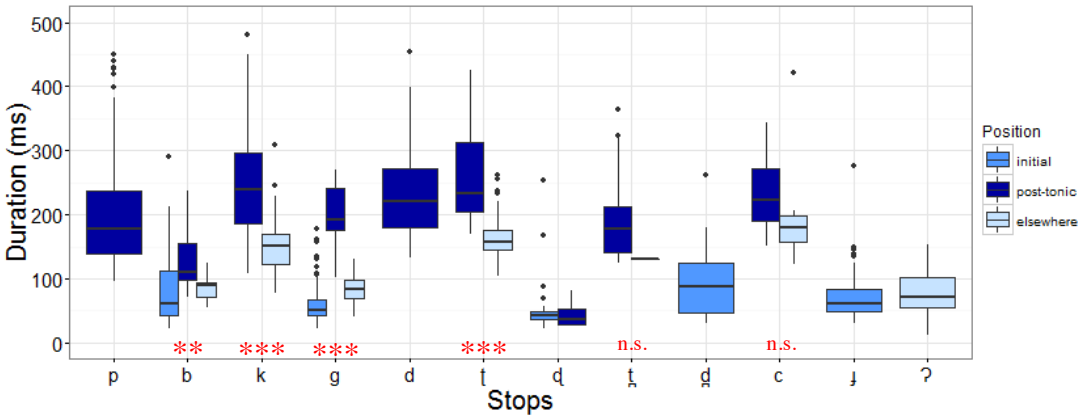


Figure 1. *Duration measurements (ms) of each stop in word initial (bright blue), post-tonic (dark blue), and intervocalic elsewhere (light blue) position. Significance indicated for relevant comparisons. (n=912)*

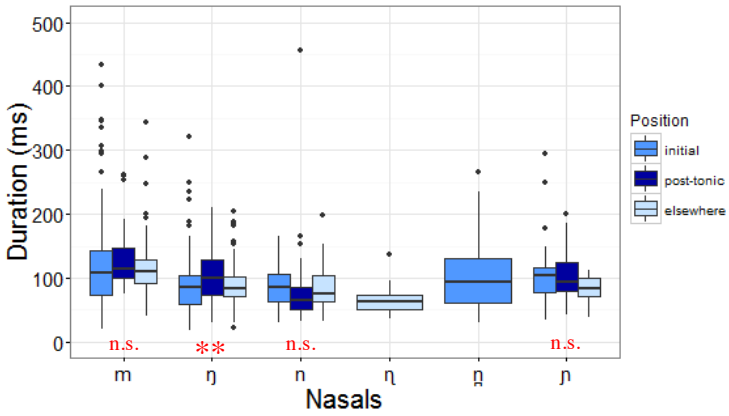


Figure 2. *Duration measurements (ms) of each nasal in word initial (bright blue), post-tonic (dark blue), and intervocalic elsewhere (light blue) position. Significance indicated for relevant comparisons. (n=1,245)*

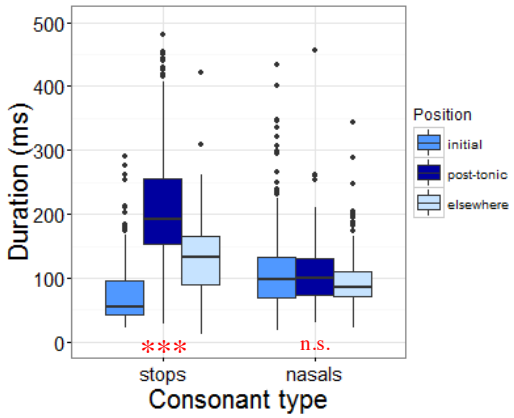


Figure 3. *Duration measurements (ms) of all stops and all nasals in word initial (bright blue), post-tonic (dark blue), and intervocalic elsewhere (light blue) position. Significance indicated for relevant comparisons. (n=2,157)*

References

[1] Baker, B. 2014. Word structure in Australian languages. In R. Nordlinger, & H. Koch (eds.), *The Languages and Linguistics of Australia*. Berlin: Mouton de Gruyter.

[2] Fletcher, J., & A. Butcher. 2014. Sound patterns of Australian languages. In R. Nordlinger, & H. Koch (eds.), *The Languages and Linguistics of Australia*. Berlin: Mouton de Gruyter.

[3] Butcher, A. 2006. Australian aboriginal languages: Consonant-salient phonologies and the "Place-of-Articulation Imperative". In J. Harrington & M. Tabain (eds.), *Speech Production: Models, phonetic processes, and techniques*. Hoboken: Taylor and Francis.

[4] Pentland, C. 2004. *Stress in Warlpiri: Stress domains and word-level prosody*. The University of Queensland, Unpublished Masters of Philosphy thesis.

[5] Fletcher, J., H. Stoakes, D. Loakes, & R. Singer. 2015. Accentual Prominence and Consonant Lengthening and Strengthening in Mawng. *Proceedings of the 18th ICPhS* (Glasgow, UK).

[6] Schiel, F., C. Draxler & J. Harrington. 2011. Phonemic segmentation and labelling using the MAUS technique. In *New tools and methods for very-large-scale phonetics research workshop program and abstracts/papers* (Philadelphia, PA).

[7] Wilkinson, Melanie. 2012. *Djambarrpuyŋu: A Yolŋu variety of northern Australia*. Muenchen: Lincom Europa

[8] Winkelmann, R., & G. Raess. 2014, 26-31 May. Introducing a Web Application for Labeling, Visualizing Speech and Correcting Derived Speech Signals. *LREC'14* (Reykjavik, Iceland).

Wednesday, 13:15 - 14:15
Poster Session 5, #14

Pragmatic factors and intonational overlap: vocatives and imperatives compared

Sergio Robles-Puente
West Virginia University

Vocative and imperative utterances have been described as indecomposable 'vocativic unitary sentences' that have the same addressing function and similar intonational properties (Jakobson, 1971; Sonnenhauser and Noel Aziz Hanna, 2013; Abuladze and Ludden, 2013). Nevertheless, no study has looked in detail at the intonations of these two kinds of sentences side by side. The aim of this paper is to examine how strong the intonational link between vocatives and imperatives is and study their potential melodic overlap in Spanish. In order to do so, the intonation and phonetic characteristics of 1540 one-word productions are analyzed together with other relevant pragmatic factors like the level of formality and the degree of insistence.

Thirty-five native speakers of Spanish produced semi-spontaneously isolated names and verbs in formal and informal settings. Half of the situations required repeating the verb or the name in order to get a production with insistence. The syllabic structure and stress patterns of names and verbs were controlled for as indicated in Table 1. A total of 840 vocatives (35 subjects x 6 vocatives x 2 formality contexts x 2 levels of insistence) and 700 imperatives (35 subjects x 10 imperatives (6 informal + 4 formal) x 2 levels of insistence) were analyzed.

Results indicate that both pragmatic contexts share multiple contours (Table 2); namely L* H%, L+H* HL%, L+H* !H%, L+H* L%, L+H* H%, L+H* LH%. As indicated in Figures 1a and 1b, not all tonal movements were equally attested across all contexts and only L+H* L% and L+H* HL% were regularly used by most subjects. Two Generalized Linear Mixed Models (GLMM) were run in order to determine if the use of tones varied depending on factors like the formality and the level of insistence of the situation. Contour, repetition, formality and all their combinations were considered fixed factors and subject a random factor. Results for imperatives indicate that there was a main effect of contour ($p < .001$) and significant interactions were also found between contour and repetition ($p < .001$), contour and formality ($p < .001$) and contour repetition and formality ($p < .022$). These were explained because L+H* L% was widely used in all productions followed by L* L% especially in formal situations. L+H* LH% and L+H* HL% were mainly used in informal contexts. The GLMM indicated that in vocatives there was a main effect of contour ($p < .001$) and significant interactions were found between contour and formality ($p < .001$) and contour repetition and formality ($p < .031$). These were the result of L+H* L% being the preferred tonal contour in vocatives; L+H* HL% and L* H% were also common in formal situations but the latter was hardly attested in informal ones. The most interesting finding is that according to both GLMMs L+H* L% was never used less than any other contour no matter what the contextual factors were. A more detailed phonetic analysis of vocatives and imperatives with this L+H* L% configuration revealed that productions in informal settings had more intensity ($p < .001$), a higher pitch on average ($p < .003$), and wider pitch excursions ($p < .004$) than those produced in the formal situations.

Current findings reveal that there is an intonational overlap between vocatives and imperatives in Spanish with up to six melodies found in both contexts. Nevertheless, L+H* L% is the most used contour regardless of the pragmatic factors, making this toneme the default melody not only for short statements but for imperatives and vocatives as well. Interestingly, speakers modify its phonetic properties of intensity and F0 depending on the formality of the situation.

formality	stress	stressed syllable	vocatives	imperatives
formal	paroxytone	CV	Ma**le**na Ma**no**lo	aban**do**ne (*give up*) remo**de**le *(remodel)*
		CVC	Do**min**go A**man**da	de**man**de *(sue)* re**ven**da *(resell)*
	oxytone	CVC	Miguel**ín** Mari**bel**	NA
informal	paroxytone	CV	Ma**le**na Ma**no**lo	or**de**na *(clean up)* ter**mi**na *(finish)*
		CVC	Do**min**go A**man**da	de**man**da *(sue)* re**ven**de *(resell)*
	oxytone	CVC	Miguel**ín** Mari**bel**	ani**mad** *(cheer up)* co**med** *(eat)*

Table 1: List of names and verbs used in the vocative and imperative situations.

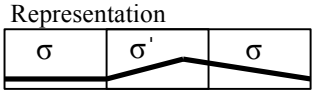
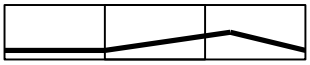
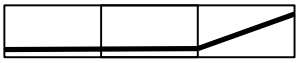
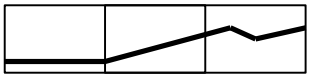
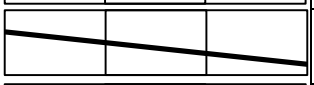
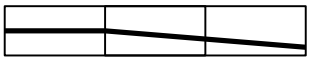
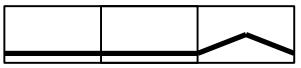
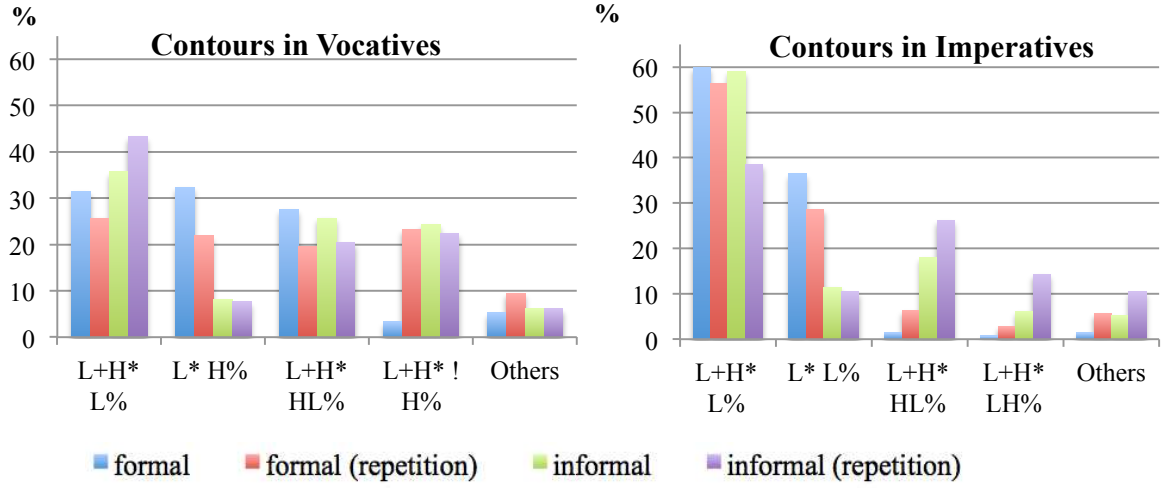
Toneme	Representation	Vocative	Imperative
L+H* L%		34% 286/840 34 subjects	53.5% 368/700 35 subjects
L+H* HL%		23.3% 196/840 28 subjects	13% 104/700 32 subjects
L+H* !H%		18.3% 154/840 27 subjects	1.1% 8/700 7 subjects
L* H%		17.5% 147/840 26 subjects	2.2% 16/700 10 subjects
L+H* H%		5.7% 48/840 13 subjects	1.1% 8/700 4 subjects
L+H* LH%		1.1% 9/840 5 subjects	6% 48/700 13 subjects
L* L%		0% 0/840 0 subjects	21.7% 137/700 28 subjects
H+L* L%		0% 0/840 0 subjects	0.3% 2/700 1 subject
L* HL%		0% 0/840 0 subjects	1.1% 9/700 4 subjects

Table 2: Tonal configurations in vocatives and imperatives.



Figures 1a and 1b: Percentage of vocative and imperative contours in different pragmatic conditions.

Is the voicing-dependant duration of obstruents physiological in French?

Yohann Meynadier, Yulia Gaydina, Antoine Giovanni
Aix Marseille Univ, CNRS, LPL, Aix-en-Provence, France

The relation between phonetic properties and phonological features has been the object of many theoretical suggestions which attempt to link the acoustic signal and linguistic representations. Phonetic redundancy and covariation in phonological contrasts are central points in this questioning. A correlation between consonant durations and voicing has been widely documented across many languages, e.g. [1]. The durational redundancy or covariation in voicing is largely seen as supported by a physiological substratum, e.g. an aerodynamic constraint on vocal vibration making voiced obstruents shorter [2, 3]. However other work, for example on French, has shown that the difference in obstruent duration according to underlying voicing feature resists complete voice assimilation [4] and total devoicing, as in whispered speech [5], arguing for linguistic conditioning. Here, the physiological vs. phonological conditioning of voicing-dependant durations of French obstruents was tested by comparing five normal and pathological phonations differing in the nature of the phonatory organ and source for voicing (see Figure 1): (i) MODal voice, i.e. periodic laryngeal phonation; (ii) WHIspered voice, i.e. non-periodic laryngeal phonation; (iii) TUCker voice due to a partial laryngectomy, i.e. non-periodic laryngeal phonation; (iv) ESOphageal voice due to a complete laryngectomy, i.e. non-periodic non-laryngeal phonation produced by an aerodynamic excitation of the esophagus; (v) Pseudo-WHIspered voice due to complete laryngectomy and no use of esophagus, i.e. non-periodic supralaryngeal voice produced by an aerodynamic excitation of only the vocal tract.

Acoustical durations were measured from 6 pairs of voiced-voiceless obstruents, i.e. /b-p/, /t-d/, /k-g/ and /f-v/, /s-z/, /ʃ-ʒ/ in initial (for fricatives), medial and word-final positions of isolated lexical words read in random ordered lists (one repetition). Table 1 reports information about speakers and data for each phonation type. Statistical effect of *Voicing* (voiced vs. voiceless) was tested by three-way ANOVAs with *Voicing*, *Articulation* (stop vs. fricative) and *Lexical Position* as the fixed effects (Table 1). The statistical comparisons between MOD and other phonations were stated on *Voicing* and *Phonation* interaction in two-ways ANOVAs included only the data for the same obstruents in the same lexical position. Reported here as a pilot study: one P-WHI speaker was analysed but not statically compared with MOD; the other one was excluded because of its total unintelligibility.

In all phonation types and word positions, the underlying voiced stops or fricatives are significantly shorter than the phonological voiceless obstruents. Table 1 shows that the voicing-dependant difference is significantly preserved regardless of phonatory organ and acoustical source type, as is confirmed by the absence of *Voicing*Phonation* interactions for every comparison with the MOD condition. To neutralize the speaking rate variation between speakers and conditions, mean ratio durations across speakers were calculated as the duration difference, i.e. voiceless consonant duration minus voiced consonant duration, divided by the voiceless consonant duration. Figure 1 shows a gradual reduction increasing with the distance from the production mechanism of modal voice used by healthy subjects. However any clear boundary seems match the change of the phonatory organ (laryngeal vs. non laryngeal) or of the phonatory source (periodic vs. non- periodic). Moreover, although phonetically reduced, the ratio of durational differences of the underlying voicing contrast remains fairly large (around 0.3). The physiological conditioning therefore seems to have only a limited effect, since the duration contrast overcomes the various physical constraints of the different phonatory mechanisms. The resistance of voicing-dependant durations to laryngectomies argues for an encoding of the systematic phonetic information at a phonological level and/or for a "phonetic knowledge" component in the grammar [6].

[1] Butcher A. (1992). Intraoral pressure as an independent parameter in oral stop contrasts. *Proceedings of the 4th Australian International Conference on Speech Science and Technology* (pp. 286–291). Brisban.
[2] Snoeren N., Seguí J., Hallé P. (2008). On the role of regular phonological variation in lexical access: Evidence from voice assimilation in French. *Cognition* 108(2), 512–521.
[3] Rothenberg M. (1968). *The Breath-Stream Dynamics of Simple-Released Plosive Production*. New York: Karger.
[4] Ohala, J. J. (1983). The origin of sound patterns in vocal tract constraints. In P. F. MacNeilage (ed.), *The Production of Speech* (pp. 189–216). New York: Springer-Verlag.
[5] Meynadier Y., Gaydina Y. (2013). Aerodynamic and durational cues of phonological voicing in whisper. *Proceedings of the 14th Interspeech*, 335-339. Lyon.
[6] Kingston J. & Diehl R. (1994). Phonetic Knowledge. *Language*, 70(3), 419-454.

Phonation	N speakers	Data type	N data	$C_{[-voi]}$ duration	$C_{[+voi]}$ duration	*Voicing* *p*
MOD	15	F: initial, median, final S: median, final	1444	203	117	*F(1,1444)=1199* *p < .001*
WHI	15	F: initial, median, final S: median and final	1218	201	127	*F(1,1218)=733.25* *p < .001*
TUC	14	S: median	423	113	78	*F(1,423)=168.721* *p < .001*
OES	6	F: initial, median, final S: median, final	1135	220	170	*F(1,1135)=26.806* *p < .001*
P-WHI	1	F: initial, median S: final	95	152	107	*F(1,95)=30.223* *p < .001*

Table 1 : Corpus information (F for fricatives and S for stops in observed word positions) and statistical effect of *Voicing* on cross-speaker mean durations (in ms) by phonation conditions.

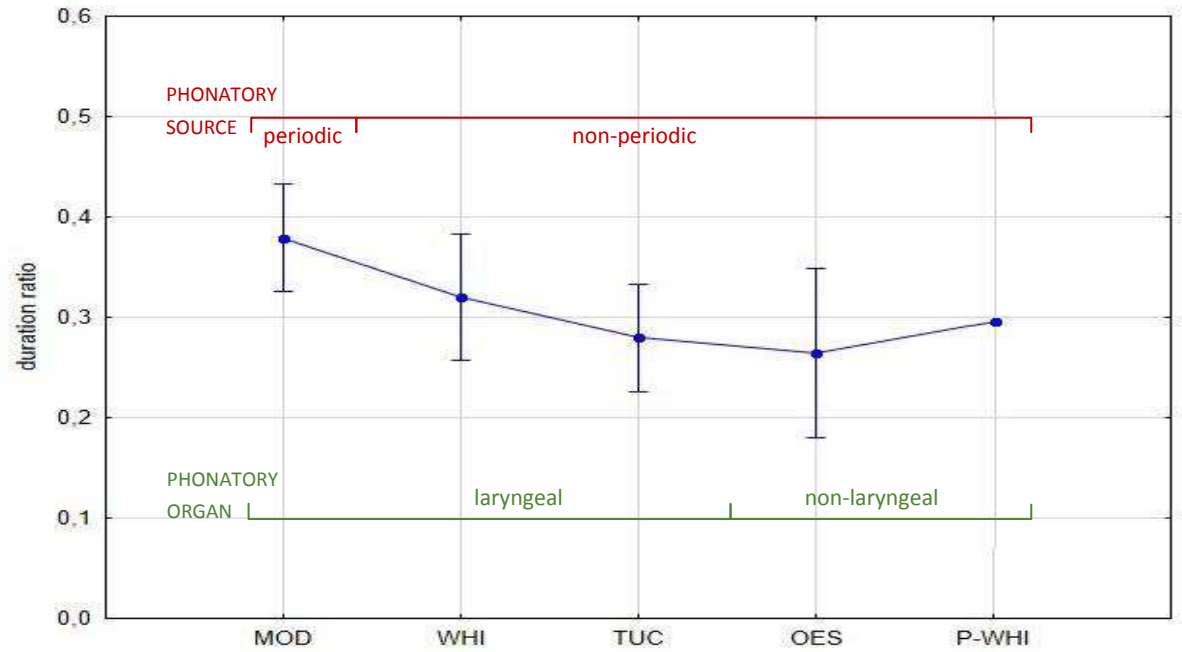


Figure 1 : Cross-speaker mean ratio of duration difference between voiced and voiceless obstruents (see text) for the five phonation conditions (bars represent standard deviation). Values higher than 0 indicate longer voiceless consonants.