Abstract booklet

PaPE

Lecce, 17-19 June 2019

Welcome to Lecce to PaPE 2019!

This is the third Phonetics and Phonology in Europe conference, following successful conferences in Cambridge, UK (June 2015) and Cologne, Germany (June 2017). Prior to 2015, a series of biennial PaPI (Phonetics and Phonology in Iberia) conferences dates back to 2003.
The idea of organising PaPE in Italy was strongly endorsed by the Italian
Association of Speech Sciences (AISV – Associazione Italiana di Scienze della Voce) – ISCA Special Interest Group.
As in the previous conferences, PAPE 2019 is an interdisciplinary forum for research on the sound properties of language as approached from multiple perspectives, with the general aim to promote discussion on the relation between phonetics and phonology, as well as the cross-fertilization of phonetics and phonology with other fields, such as psycholinguistics, computational linguistics, neurosciences and speech technology.
In this vein, PaPE 2019 continues the tradition of the series in welcoming submissions on any aspect of phonetics and phonology, especially from an interdisciplinary perspective. This year we particularly encourage papers dealing with all aspects of language variation. Contributions are therefore welcome on language variation in all possible domains, such as, for example, sociophonetics, language learning and development, speech pathology, speech accommodation and multimodal communication.
For PaPE 2019 we received 192 submissions. Of these, 104 will be presented at the conference (26 oral presentation, 78 poster presentation.
We would like to warmly thank the over 200 reviewers involved in the revision process!!!

Looking forward to your presentations and to stimulating discussions inside and outside the conference sessions!

PaPE 2019 Organising Committee

Barbara Gili Fivela
Cinzia Avesani
Michelina Savino

Local organisation

Sonia d'Apolito
Paolo Lorusso
Mirko Grimaldi
Francesco Sigona

Sponsors

A grateful thank to our sponsors, the Association for Laboratory Phonology (LabPhon), Fondazione Puglia and the Associazione Italiana di Scienze della Voce (AISV)

Phonetics and Phonology in Europe 2019

Lecce, 17-19 June 2019
Bari, 20 June 2019

FINAL PROGRAM

SUNDAY 16 June	
18:00 -20:00	Registration

MONDAY 17 June	
08:00 – 09:00	Registration
09:00 - 09:30	**Opening ceremony**
09:30 – 10:30	**Keynote speech:** Janet Pierrehumbert, *Regular sound change through the lens of exemplar theory*
10:30 – 11:00	Coffee break
	Oral session 1: From short term accommodation to change I chair Pilar Prieto
11:00 – 11:30	Stephan Schmid, Markus Jochim, Nicola Klingler, Michael Pucher, Urban Zihlmann and Felicitas Kleber *Vowel and consonant quantity in southern German varieties: typology, variation, and change*
11:30 – 12:00	Mary Baltazani, Joanna Przedlacka and John Coleman *Intonation in contact: Athenian, Cretan, Corfiot and Venetian declaratives*
12:00 – 12:30	Anne Cutler, Laura Ann Burchfeld and Mark Antoniou *Language-specificity and experience-dependence of phonetic adjustment to talkers*
12:30 – 13:00	Jenny Yu and Katharina Zahner *Compensation strategies in non-native English and German productions: Evidence for prosodic transfer and adjustment*
13:00 – 14:00	Lunch break

14:00 -15:00	**Poster session 1** chair: Stephan Schmid
	Oral session 2: From short term accommodation to change II chair: Janet Fletcher
15:00 – 15:30	Wenling Cao *Short-term accommodation of Hong Kong English towards RP and GenAmE*
15:30 – 16:00	Christiane Ulbrich *Phonetic accommodation on the segmental and the suprasegmental level of speech in native-non-native collaborative tasks*
16:00 – 16:30	Coffee break
16:30 – 17:30	**Poster session 2** chair: Francesco Cangemi
	Oral session 3: Factors affecting perception I chair Anne Cutler
17:30 – 18:00	Laurence White, Silvia Benavides-Varela, Katalin Mády and Sven Mattys *The primary importance of onsets: Timing and prediction in speech segmentation*
18:00 – 18:30	Chelsea Sanker *Effects of coda laryngeal features and stimulus language on perceived vowel duration*
18:30 - 19:00	Nadja Kerschhofer-Puhalo *"Similar or different?" – The phonetics and phonology of similarity in non-native vowel perception*

TUESDAY 18 June	
09:00 – 10:00	**Keynote speech:** Wolfran Ziegler, *The challenge of word articulation: A neurophonetic view*
	Oral session 4: Pathological speech chair Giovanna Marotta
10:00 – 10:30	Daria D'Alessandro, Michaela Pernon, Cécile Fougeron and Marina Laganaro *Anticipatory V-to-V coarticulation in French in different Motor Speech Disorders*
10:30 – 11:00	Caterina Petrone, Melody Zira, Christelle Zielinski and Elisa Sneed German *Cognitive abilities and prosody in question-response interactions: A clinical study*
11:00 – 11:30	Coffee break
	Oral session 5: Variation and modelling chair Martine Grice
11:30 – 12:00	Georg Lohfink, Argyro Katsika and Amalia Arvaniti *Variation, variability and category overlap in intonation*
12:00 – 12:30	Leonardo Lancia and Cristel Portes *The contribution of dynamics to the perception of tonal alignment*

12:30 – 13:00	Simon Roessig and Doris Mücke *Prosodic variation in a multi-dimensional dynamic system*
13:00 – 14:00	Lunch break
14:00 -15:00	**Poster session 3** chair: Stefan Baumann
	Oral session 6: Factors affecting perception II chair: Mariapaola D'Imperio
15:00 – 15:30	Caterina Ventura, Martine Grice, Michelina Savino, Aviad Albert and Petra Schumacher *Perceptual evaluation of post-focal prominence in Italian by L1 and L2 naïve listeners*
15:30 – 16:00	Gilbert Ambrazaitis and David House *Multimodality in prominence production and its sensitivity for lexical prosody*
16:00 – 16:30	Coffee break
16:30 – 17:30	**Poster session 4** chair: Maria del Mar Vanrell
	Oral session 7: L2 learning chair Aoju Chen
17:30 – 18:00	Fabian Santiago and Paolo Mairano *Prosodic effects on L2 French vowels: a corpus-based investigation*
18:00 – 18:30	Peng Li, Florence Baills and Pilar Prieto *Durational hand gestures facilitate the learning of L2 vowel length contrasts*
19:15	**Social dinner**

WEDNESDAY 19 June	
09:00 – 10:00	**Keynote Speech:** Sara Bögels**,** *Turn-taking: Early planning and late cues*
	Oral session 8: Processing of information chair Amalia Arvaniti
10:00 – 10:30	Stefan Baumann and Petra Schumacher *The incremental processing of pitch accents, information status and focus*
10:30 – 11:00	Mikael Roll, Pelle Söderström and Merle Horne *Pre-activation negativity (PrAN): A neural index of predictive strength of phonological cues*
11:00 – 11:30	Coffee break
	Oral session 9: Sign language chair Louis Goldstein
11:30 – 12:00	Carlo Geraci, Justine Mertz, Jessica Lettieri, Shi Yu and Natasha Abner *Typological and Historical relations across sign languages The view from articulatory features*

12:00 – 12:30	Valentina Aristodemo, Chiara Annucci, Justine Mertz, Giustolisi Beastrice, Carlo Geraci and Caterina Donati *Measuring phonological complexity in Sign Language*
12:30 – 13:00	Adelaide Silva and André Xavier *Libras and Articulatory Phonology*
13:00 – 14:00	Lunch break
14:00 – 16:00	**Poster session 5** chair: Caterina Petrone
16:00 – 16:30	Coffee break
	Oral session 10: Speech production chair Cinzia Avesani
16:30 - 17:00	Sarah Harper *Intrasegmental gestural timing for American English /ɹ/ in isolated and connected speech*
17:00 -1 7:30	Miran Oh, Dani Byrd, Louis Goldstein and Shrikanth Narayanan *Vertical larynx actions and larynx-oral timing in ejectives and implosives*
17:30 - 18:00	Christopher Carignan *Network analytics reveals patterns in many-to-one articulatory-to-acoustic strategy*
18:00	**Closing ceremony**

THURSDAY 20 June	
10:30-18:30	**Satellite Workshop in Bari** “Prominence between Cognitive Functions and Linguistic Structures” (COFLIS)

Posters

Poster Session 1 – Monday 17 June 14:00 -15:00		
1	Jolanta Sypiańska	*The L1 of Polish expats in Denmark – a comparison of three emigration waves*
2	Janet Fletcher, Rosey Billington and Nick Thieberger	*Word order, focus, and prosodic variation in Nafsan (Vanuatu)*
3	Jennifer Sander, Barbara Höhle and Aude Noiray	*From the eye to the mouth: does a developmental shift in attention co-emerge with the emergence of babbling?*
4	Marcel Schlechtweg and Holden Härtl	*Quotation marks are expressed in spoken language*
5	Melina Heinrichs, Marcel Schlechtweg and Marcel Linnenkohl	*Acoustic differences between homophonous singular-plural nouns in German*
6	Michela Russo	*Clusters and complex onsets in Romagnolo. A cross-linguistically syllabification algorithm?*
7	Anisia Popescu, Lisa Hintermeier, Stella Kruger and Aude Noiray	*Does the acquisition of reading affect speech production?*
8	Cristel Portes and Arndt Riester	*How prominence shift and phrasing combine to mark focus in French*
9	Karolina Bros, Martin Meyer and Volker Dellwo	*Default and exceptional stress processing in Spanish: an ERP analysis*
10	Mark Gibson, Ferenc Bunta, Charles Johnson and Miriam Huarriz	*Using Random Forests to compare production patterns by bilingual English-Spanish speaking children with cochlear implants and their peers with normal hearing*
11	Jane Mertens, Anne Hermes and Doris Mücke	*Velocity profiles and prosodic variation in younger and older speakers*
12	Peter Staroverov and Sören Tebay	*Allophony and history of Mee velar lateral*
13	Giovanna Lenoci, Chiara Celata and Irene Ricci	*Articulatory-to-acoustic relations in Childhood Apraxia of Speech: Vowel Production*

Poster Session 2 - Monday 17 June 16:30 -17:30		
1	Nicolas Audibert and Cecile Fougeron	*Are people as old as they sound? Acoustic, regional and generational effects*
2	Maria Del Mar Vanrell and Ingo Feldhausen.	*The realization of focus in monolingual and bilingual Spanish*
3	Christine T. Röhr, Stefan Baumann, Martine Grice and Petra Schumacher.	*Signal- and expectation-based processing of pitch accent types*
4	Elisabeth Delais-Roussarie and Claire Beyssade	*Information seeking vs. rhetorical questions: from gradience to categoricity*

5	Patrick L. Rohrer and Elisabeth Delais-Roussarie	*Beat Gestures as Prosodic Domain markers in French: A case study*
6	Nicole Dehé and Bettina Braun	*Icelandic question intonation*
7	Monika Konert-Panek	*The significance of frequency effects in stylisation patterns. LOT unrounding and PRICE monophthongisation in popular music singing accent*
8	Ingrid Vilà-Giménez and Pilar Prieto	*Producing rhythmic beat gestures while retelling a story: positive effects of a gesture-based training session on children's narrative performance*
9	Brechtje Post and Samantha Jones	*Similarity and contrast in L1 pronunciation attrition in bilinguals*
10	Roxane Bertrand, Corine Astésano and Noël Nguyen	*Prominence and boundary are two distinct phenomena in French: perceptual evidence*
11	Haoru Zhang and Chelsea Sanker	*Phonetic Convergence in Mandarin*
12	Motoko Ueyama, Ryoko Hayashi and Aaron Lee Albin	*Development of an L2 Japanese speech corpus for the comparison of prosody across diverse L1 groups*
13	Carlo Geraci and Justine Mertz	*Theory-Description-Theory: A round trip in French sign language phonology*

Poster Session 3 - Tuesday 18 June 14:00 – 15:00		
1	Márcia Cristina Do Carmo	*Variable vowel raising: acoustic analysis of word-initial pretonic mid-vowel /e/ in Brazilian Portuguese*
2	Martina Sciuto, Giulano Bocci and Vincenzo Moscati	*Clause typing and edge tones in early Italian: a longitudinal study*
3	Riccardo Orrico, Renata Savy and Mariapaola D'Imperio	*Variability in the perception of epistemic valence in Salerno Italian question tunes*
4	Stella Krüger and Aude Noiray	*How much in advance can listeners perceive upcoming speech targets? Insights from children and adults*
5	Yifan Yang, Rachel Walker and Alessandro Vietti	*Variation of sibilants in three Ladin dialects*
6	San-Hei Kenny Luk and Daniel Pape	*Perception of Canadian English sibilants: Processing of acoustic information or underlying (articulatory) vocal tract configurations?*
7	Paolo Mairano and Fabian Santiago	*Do Italian native learners produce and expect geminate consonants in L2 French?*
8	Elisa Pellegrino, Hanna Ruch, Thayabaran Kathiresan and Volker Dellwo	*Rhythmic Convergence and Divergence in two Swiss German dialects*
9	Joe Rodd, Hans Rutger Bosker, Mirjam Ernestus, Antje Meyer and Louis Ten Bosch	*The speech production system is reconfigured to change speaking rate*
10	Luma da Silva Miranda, João Antônio de Moraes and Albert Rilliard	*The perception of prosodic cues in Brazilian Portuguese statements and echo-questions: analysis by resynthesis*
11	Nicholas Nese	*Accommodation processes in the speech of university college students: a real-time sociophonetic approach*
12	Yaru Wu, Martine Adda-Decker, Cécile Fougeron and Cédric Gendrot	*How do French Cʁ# cluster realizations vary across speaking style?*
13	Bettina Braun and María Biezma	*Prenuclear L*+H but not L+H* leads to the activation of alternatives in German*

Poster Session 4 - Tuesday 18 June 16:30 – 17:30		
1	Anna Huszár	*The prosody of impersonating characters in storytelling speech*
2	Oana Niculescu, Ioana Vasilescu and Martine Adda-Decker	*Phonetic encoding of phonological representation of hiatus in Romanian: a study of durational patterns*
3	Glenda Gurrado	*On the prosodic cues of ironic exclamatives. A pilot study*
4	Eduardo García-Fernández	*Tune choice in utterance-initial vocatives in Asturian*
5	Patrizia Sorianello	*A prosodic study of rhetorical questions in Italian*
6	Andrea Deme, Márton Bartók, Tekla Etelka Gráczi, Tamás Gábor Csapó and Alexandra Markó	*The effect of pitch accent on V-to-V coarticulation induced variability of vowels*
7	Alexei Kochetov and Kiranpreet Nara	*Gestural adaptation when 'broadening' an L2 accent: An exploratory EPG study*
8	Nuria Martínez García, Francesco Cangemi and Martine Grice	*Resistance to resyllabification in Yucatecan Spanish: effects of prosodic structure and contact dynamics*
9	Lorenzo Spreafico, Alessandro Vietti, Alessia Pini and Simone Vantini	*Investigating articulatory variability across recording sessions using a functional clustering technique*
10	Janina Mołczanow, Beata Łukaszewicz and Anna Łukaszewicz	*Disentangling metrical prominence from segmental and word-boundary effects*
11	Alexandra Markó, Márton Bartók, Tamás Gábor Csapó, Andrea Deme and Tekla Etelka Gráczi	*Variability in realization of focal accent in Hungarian – articulatory and acoustic data*
12	Priscila Santos, Sónia Frota and Marisa Cruz	*The yes-no question contour in Northern Brazilian Portuguese: revisiting the geographical continuum*
13	Darya Kavitskaya and Sharon Inkelas	*Cluster simplification in Russian children with Specific Language Impairment*
14	Ottavia Tordini, Vincenzo Galatà, Cinzia Avesani and Mario Vayra	*Acoustic evidence of /s/-retraction in the Northern Veneto dialect: a case-study*

Poster Session 5 - Wednesday 19 June 14:00 – 16:00		
1	Anna Kohári, Katalin Mády, Andrea Deme, Uwe Reichel and Ádám Szalontai	*Utterance-final lengthening in infant-directed speech*
2	Rachida Ganga, Haoyan Ge, Marijn Struiksma, Virginia Yip and Aoju Chen	*Processing prosodic information in sentences with "only" in a second language*
3	Mayuki Matsui and Silke Hamann	*Voicing, devoicing or contrast enhancement? Russian homorganic nasal-stop sequences in a devoicing context*
4	Corine Astésano	*The prosodic word as the domain of French accentuation - Empirical evidence*
5	Colleen Fitzgerald, Rebecca Ebert and Jason Whitfield	*Acoustic Differences among English -s Allomorphs in a Children's Book Reading Task*
6	Amandine Michelas and Maud Champagne-Lavau	*To what extent the French prosodic encoding of contrast is addressee-oriented?*

7	Marina Vigário, Marisa Cruz and Sónia Frota	*Why tune or text? Explaining crosslinguistic variation in the resolution of tune-text conflicts*
8	Nuria Esteve-Gibert, Carme Muñoz, Natàlia Fullana, Ingrid Mora-Plaza, Lena Vasylets and Joan Carles Mora	*Learning non-native phonological contrasts through multimodal input: the contribution of the tactile sense*
9	Conceição Cunha, Samuel Silva, Arun Joseph, António Teixeira, Catarina Oliveira, Paula Martins and Jens Frahm	*Variability in the dynamic of nasal vowels in European Portuguese*
10	Kiwako Ito, James German, Elisa Sneed German and Caterina Petrone	*Intonation affect perlocutionary meaning in requests and offers*
11	Tekla Etelka Gráczi, Mária Gósy, Valéria Krepsz, Anna Huszár, Nóra Damásdi, Alexandra Markó and Ákos Gocsál	*Voice patterns associated with age and gender of speakers across the lifespan*
12	Mariia Pronina, Iris Hübsch, Ingrid Vilà-Giménez and Pilar Prieto	*Assessing pragmatic prosody in 3- to 4-year-old children*
13	Sumio Kobayashi and Amalia Arvaniti	*A corpus-based analysis of Japanese rhythm and mora duration*
14	Andressa Toni	*[ˈpɾa.tʊ], [ˈpa.tʊ], [ˈpla.tʊ], [ˈpɐɾa.tʊ], [ˈpa.tɾʊ], [ˈpaɾ.tʊ]: variable outputs for CCV syllables in the acquisition of Brazilian Portuguese*
15	Alba Aguete Cajiao and Elisa Fernández Rei	*Vowel height variation due to prosodic strengthening*
16	Paula Orzechowska	*Conflicting forces of place and manner of articulation: A reaction times study on Polish phonotactics*
17	Marta Ramon-Casas, Natasha Sanz Gauntlett, Ferran Pons and Laura Bosch	*Audiovisual speech and the discrimination of a native fricative contrast: adult and infant data*
18	Sejin Oh, Jason Bishop and Chen Zhou	*Effects of L2 Experience on the Realization of L1 Phonological Neutralization: Incomplete Devoicing in Bulgarian-English Bilinguals*
19	Nicholas Henriksen, Andries Coetzee, Lorenzo Garcia-Amaya, Jiseung Kim and Daan Wissing	*Uncovering the geographic origin of immigrant communities: Transitional gliding in Patagonian Afrikaans*
20	Olga Dmitrieva, Amy Hutchinson, Allard Jongman, Joan Sereno and Alexis Tews	*The effect of instructed second language learning on the acoustic properties of first language speech*
21	Mairym Llorens Monteserin	*Phonetics and distribution of Tourette's verbal tics produced during active speech*
22	Mariapaola D'Imperio and Lidiia Dorokhova	*Rise shape dynamics is sufficient to distinguish question and continuation rises in French*
23	Yuki Asano, Anne Cutler, Andrea Weber and Ann-Kathrin Grohe	*Experience with an uptalk variety and perception of high rising terminal contours*
24	Asta Kazlauskiene and Sigita Dereskeviciute	*The Intonational Patterns of Interrogative Sentences in Lithuanian*
25	Adèle Jatteau, Ioana Vasilescu, Lori Lamel and Martine Adda-Decker	*Final devoicing in the "pool of variation": A large-scale corpora approach with automatic alignment*

Keynote speech I

Regular sound change through the lens of exemplar theory

Janet B Pierrehumbert

Dept. of Engineering Science, University of Oxford

The classic theory of sound change distinguishes regular sound changes from analogical changes. Regular sound changes gradually transform the phonetic realisations of phonemes, and have been claimed by many scholars to affect all words simultaneously and equally. Analogical changes are categorical changes in the expressions of morphosyntactic categories, typically bringing rare words into conformity with more common words. This strong distinction rests on strong assumptions about modularity in linguistic systems. These assumptions have given way in the light of detailed experimental studies showing that the cognitive representations of individual words include remarkably detailed phonetic and contexual information. Exemplar theory provides a formal framework for analysing such effects.

This talk will present the empirical evidence for word-specific effects in regular sound change, and also present an exemplar model that can generate them. The model builds on experimental results about the effects of word frequency and typicality in perceiving speech and laying down memory traces. I will argue that success in explaining the range of diachronic patterns observed depends crucially on adopting a hybrid approach, in which lexical, phonological, and phonetic levels all play a role.

Oral session 1

From short term accommodation to change I

Vowel and consonant quantity in southern German varieties: typology, variation, and change

Stephan Schmid[1], Markus Jochim[2], Nicola Klingler[3], Michael Pucher[3], Urban Zihlmann[1], Felicitas Kleber[2]

[1]*Phonetisches Laboratorium, Universität Zürich,* [2]*Institut für Phonetik und Sprachverarbeitung, Ludwig-Maximilians-Universität München,* [3]*Institut für Schallforschung, Österreichische Akademie der Wissenschaften*

Southern German varieties present a rather high degree of variation with respect to vowel and consonant quantity, both from the typological point of view and from the perspective of language contact and sound change. For instance, Standard German has distinctive vowel quantity (e.g. /ban/ 'ban(n.)' ~ /baːn/ 'railway') as well as a fortis-lenis contrast between stops (e.g., /ˈza̯ɪtə/ 'side' ~ /ˈza̯ɪdə/ 'silk' [1]); phonetically, the contrast between fortis and lenis stops is mainly implemented through a longer VOT (aspiration) and closure duration of the former. The phonotactics of Standard German allows for all four combinations of short and long vowels with fortis and lenis stops: VːC, VːCː, VC, VCː (for the sake of simplicity, we symbolise fortis stops as long consonants Cː). Instead, traditional Central Bavarian (CB) has a phonotactic restriction that prohibits the combinations of long vowels + fortis stops and of short vowels + lenis stops, but there are indices for an incipient sound change due to the contact with the standard variety [2, 3]. Finally, Swiss German dialects exploit vowel quantity in a systematic manner [4], but – differently from Standard German – the fortis-lenis contrast is implemented in stops solely through a difference in closure duration [5].

This study presents for the first time acoustic evidence from a cross-linguistic study for the typological diversity in three Southern German varieties from Austria, Germany and Switzerland and further evidence for a sound change currently in progress in Bavarian dialects. A total of 104 speakers recorded in Munich (12 Standard German and 24 West Central Bavarian dialect, WCB), Vienna (19 standard variety and 13 East Central Bavarian, ECB) and Zurich (20 Alemannic dialect) read standard or dialectal target words (depending on the speaker group) containing the four phonotactic combinations VːC, VːCː, VC, VCː (e.g. Standard German *wieder* 'again', *Bieter* 'tenderer', *Widder* 'ram', *bitter* 'bitter') embedded in standard or dialect sentences. Adopting an apparent-time approach [6], we included speakers who were under 30 and over 50 years old. As a measure of comparison, we calculated the ratio of the vowel duration to the duration of the vowel+stop closure sequence V/(V+C) [2]. Our hypothesis was that we would find three types of V/(V+C) ratios in German standard varieties and in the Swiss dialects (i.e. low values for VCː, high values for VːC, and intermediate values signalling equal vowel and closure proportions for VC and VːCː combination), whereas CB would show a mixed pattern according to the regional background and the two age groups.

The results show indeed that speakers of Standard German in Munich and Vienna display three rather neatly distinct types of duration ratios (i) VːC, ii) VːCː and VC, iii) VCː); the same observation holds for Zurich dialect speakers (Fig. 1). On the other hand, the two groups of Bavarian dialect speakers show a rather instable system which is moving away from the traditional phonotactic constraint described above. While Viennese dialect speakers (Fig. 2, right) show vowel shortening in the VːCː category as well as vowel lengthening in the VC category, WCB speakers' (Fig. 2, left) ratios especially of VC sequences overlap with the other three categories which is a result of generational differences in this group: whereas older speakers adhere to the traditional phonotactic constraint (resulting in only two major duration patterns), among the younger speakers a new category VC emerges (cf. Fig. 3). No such generation effects were found in the other three groups. Bimodal distributions of some of the within-category ratios can be related to word-specific pronunciations suggesting a change in terms of lexical diffusion [8].

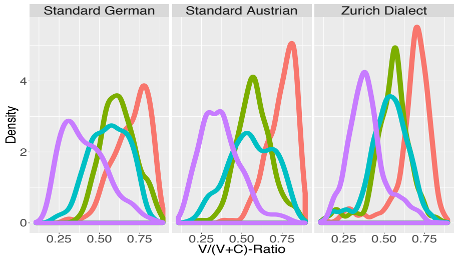


Figure 1. *Density plots of the V/(V+C) ratios in VC: (purple), VC (turquoise), V:C: (green), V:C (red) sequences separately for the three regional groups across age groups.*

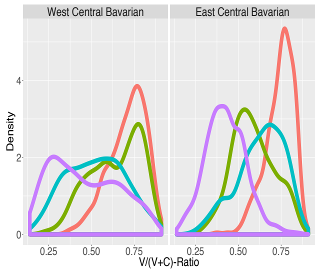


Figure 2. *Density plots of the V/(V+C) ratios in VC: (purple), VC (turquoise), V:C: (green), V:C (red) sequences in West (left) and East (right) Central Bavarian speakers.*

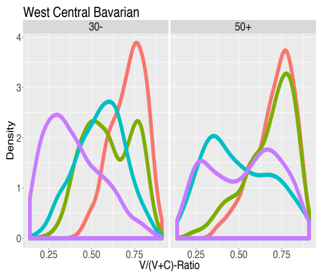


Figure 3. *Density plots of the V/(V+C) ratios in VC: (purple), VC (turquoise), V:C: (green), V:C (red) sequences in younger (left) and older (right) West Central Bavarian speakers.*

[1] Kohler, K. 1995. *Einführung in die Phonetik des Deutschen.* 2nd edition. Berlin: Erich Schmidt Verlag.

[2] Kleber, F. 2017. Complementary length in vowel-consonant sequences: Acoustic and perceptual evidence for a sound change in progress in Bavarian German. *Journal of the International Phonetic Association* 47, 1-22.

[3] Moosmüller, S., & Brandstätter, J. 2014. Phonotactic information in the temporal organization of Standard Austrian German and the Viennese dialect. *Language Sciences* 46, 84-95

[4] Fleischer, J., & Schmid, S. 2006. Zurich German. *Journal of the International Phonetic Association* 36, 243-253.

[5] Ladd, D. Robert, & Schmid, S. 2018. Obstruent voicing effects on F0, but without voicing: Phonetic correlates of Swiss German lenis, fortis, and aspirated stops. *Journal of Phonetics* 71, 229-248.

[6] Labov, W. 1963. The social motivation of sound change. *Word* 19, 273-309.

[7] Trudgill, P. 2008. Colonial dialect contact in the history of European languages: On the irrelevance of identity in new-dialect formation. *Language in Society* 37, 241-280.

[8] Labov, W. 2007. Transmission and diffusion. *Language* 83, 344-387.

Intonation in contact: Athenian, Cretan, Corfiot and Venetian declaratives

Mary Baltazani, Joanna Przedlacka, John Coleman
University of Oxford

Little work exists on the impact of long-term historical language contact on intonational features. We examine the declarative tunes in Cretan and Corfiot Greek, two varieties with a history of long-term contact with speakers of the Venetian dialect of Italian. We compare the declarative in these two varieties to the corresponding tune in Athenian Greek and Venetian. We hypothesise that, given the history of Venetian occupation of the respective islands for more than four centuries, Cretan and Corfiot will pattern with Venetian, not Athenian.

Drawn from pre-existing corpora, our data comprises spontaneous and semi-spontaneous speech, in contrast to the controlled laboratory speech analysed in the majority of intonational studies. The preliminary results reported below are based on three of the four varieties to be presented: recorded between 2001 and 2015, they comprise 157 Athenian utterances (5M speakers, mean age 50; 5F speakers, mean age 53), 120 Cretan (5M, 54; 5F, 54.4), and 44 Venetian ones (1M, 21; 1F, 19). This is work in progress and more recordings per language variety will be investigated. Standard Autosegmental–Metrical (AM) phonological analysis [1, 2] was combined with statistical modelling of f_0 curves (Functional Data Analysis, FDA), to examine the nuclear part of declarative utterances. Figure 1 shows examples of declaratives in Athenian, Cretan and Venetian as well as the region of interest (RoI) to be modelled, defined through the AM analysis and delimited by phonological landmarks P, the pre-stress vowel and V, the nuclear vowel. The RoI f_0 curves were modelled as 4th-order orthogonal (Legendre) polynomials [3], to capture both phonological and continuously varying phonetic properties of the f0 curves which have been reported to contribute to the utterance interpretation [4, 5]. Statistical significance of differences between varieties was tested using t-tests.

Preliminary results show that while declaratives in all three varieties end in a fall, there are important differences in the timing and scaling of the fall. Figure 1 shows that the f_0 fall in Athenian starts during the stressed vowel and falls faster than in Venetian and Cretan where it starts earlier and is more gradual, forming a plateau from about the middle of the stressed vowel to the end of the utterance. This impressionistic description is confirmed by the statistical analysis of the modelled curves. The slope of the Athenian fall (Figure 2 top left), reflected by the larger negative coefficient of the linear term of the polynomial (μ: -5.7; σ: 13.4), is significantly steeper than in Venetian (μ: -.29; σ: 13.4; $p < .01$), while the slopes of the Cretan fall (μ: -3.7; σ: 29.2) are not significantly different from those of the other varieties (Ath/Cretan, p=.452; Ven/Cretan, p=.451), due to their large variance. The quadratic coefficient for the Cretan curves (μ: 1.1) is significantly larger than in Athenian (μ: -14.1; $p < .001$) but not significantly different from Venetian (μ: .52; p = .955; Figure 2 top right). The sign of the mean quadratic coefficient shows that the Athenian f_0 curves steeply downwards, i.e. forms a concave downwards peak, while the Cretan f_0 curves gently upwards, i.e., it falls from a peak to form a trough (cf. Figure 1).

The AM and FDA analyses converge on a consistent picture, promising less labour intensive/costly and more reliable analyses of large datasets. These results confirm our hypothesis that the shape and scaling details of Cretan tunes pattern more like Venetian than Athenian. They suggest that declaratives in Cretan resemble Venetian both in the nuclear shape and the location of the peak. In Venetian and Cretan there is a fall that starts in the prenuclear syllable and a L pitch aligned with the stressed syllable (tentatively analysed as H+L*), as opposed to Athenian, where there is an H on the stressed syllable (H*) followed by a fall. This work showcases a promising direction for investigating the effect of language contact on regional intonational variation using natural linguistic data by combining the insights of the AM framework with a data-driven statistical analysis of f_0 curves.

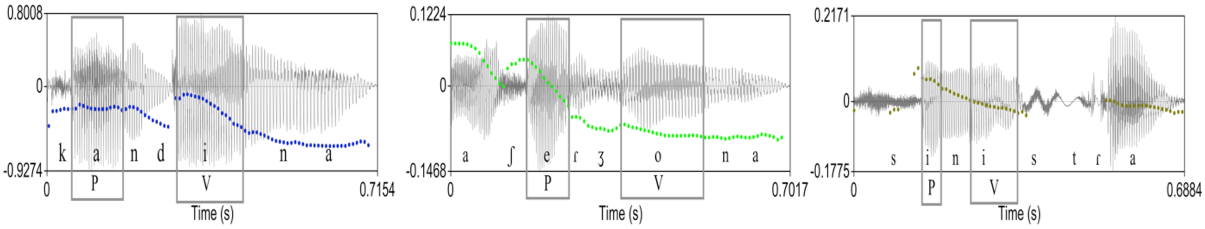


Figure 1. *Illustrations of declaratives (mean contours re the quadratic coefficient) in SMG (a), CrG (b) and Venetian (c); p = pre-stress vowel; v = last stressed vowel.*

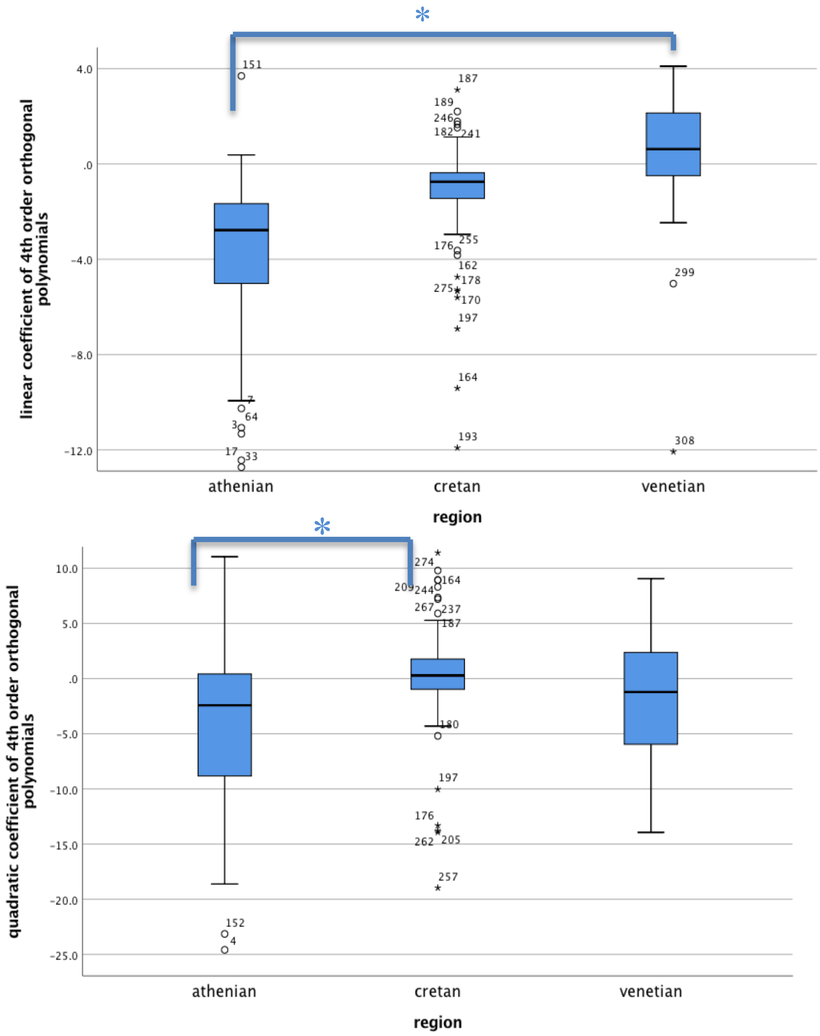


Figure 2. *Shape characteristics of the f_0 curves based on the linear (top) and quadratic (bottom) coefficients of the polynomial (see text for details).*

[1] Pierrehumbert, J. B. 1980. *The phonology and phonetics of English intonation.* PhD dissertation, MIT.

[2] Ladd, D. R. 2008. *Intonational Phonology*. 2nd edition. Cambridge: CUP.

[3] Grabe, E., Kochanski, G. & Coleman, J. 2007. Connecting intonation labels to mathematical descriptions of fundamental frequency. *Language and Speech* 50(3), 281–310.

[4] Cangemi, F. 2014. *Prosodic detail in Neapolitan Italian*. Language Science Press.

[5] D'Imperio, M. & Cangemi, F. 2009. The interplay between tonal alignment and rise shape in the perception of two Neapolitan rising accents. *Phonetics and Phonology in Iberia* (PaPI) 2009.

Language-specificity and experience-dependence of phonetic adjustment to talkers

Anne Cutler, Laura Ann Burchfield, and Mark Antoniou
The MARCS Institute for Brain, Behaviour and Development, Western Sydney University.

Human listeners adapt to newly-encountered talkers with impressive rapidity. In the past 15 years, this process has been intensively examined using a paradigm in which listeners hear ambiguous sounds which can be disambiguated by invoking existing lexical knowledge [1]. In such studies, a form midway between /f/ and /s/, for example, will be learned as /f/ if heard in words like *giraffe*, as /s/ if heard in words like *horse*, but will remain ambiguous if heard in nonwords such as *liff* or *liss*. The learning generalizes to other words with the same phoneme, enabling adaptation to new talkers on first encounter, and is stable across position in exposure words. Crucial for the present work is that such learning has been attested in both European and non-European languages, and can be successfully applied in a second languages: L2 [2-9].

In one such study, successful L2 adaptation was accompanied, unexpectedly, by failure of the same listeners (regular users of both languages) to adapt in the native language: L1 [7]. The participants were long-term emigres (averaging 22 years in the L2 environment, but still using the L1 if only with family). This finding could reflect the participants' higher age than in other L2 studies, or some aspect of the emigre situation; or, interestingly, it could indicate that talker adaptation mechanisms in a language need regular practice for optimal operation, and talking with long-known family does not require adaptation, whereby no practice occurs.

We here test this latter explanation in another group of bilinguals who are younger and not emigres, but may potentially have a similarly asymmetric number of conversational partners in their two languages: heritage language users. Again, both L1 and L2 listening were tested in the same participants; these were 24 users of Mandarin and English, average age 22.2 years, with Mandarin as language of the family and English as language of the environment.

The English and Mandarin stimulus sets were balanced for phonological acceptability in each language, and both sets had successfully induced perceptual learning in other listeners. For each language, pre-tests established a maximally ambiguous percentual mix of /f/ and /s/, which was then substituted for either /f/ or /s/ in word-medial position in exposure-phase words such as *traffic* vs. *gossip*, or *bu4fa3* 'illegal' vs. *kuan1song1* 'loose'. All participants took part in both English and Mandarin test sessions, at least 14 days apart; language order was counter-balanced. In each session, participants first performed a lexical decision task as training phase, in which the ambiguous sound replaced either /f/ (inducing learning that this sound should be interpreted as /f/) or /s/ (learning it to be /s/). They then carried out a phonetic categorization task, in which they heard 5 tokens along a continuum from a near-endpoint /f/ to a near-endpoint /s/, each 30 times, in randomized order. Perceptual learning would be revealed by expansion of the trained /f/ or /s/ category. Each participant was trained on one sound per language; the trained category was the same across languages for half the group, different for the other half. At study outset, participants filled in an extensive language use questionnaire.

Figure 1 shows the results of the phonetic categorization experiment in each language. A significant perceptual learning effect can be seen with the English materials, $F(1, 22) = 4.47$, $p < .05$, but with the Mandarin stimuli, the same listeners showed only very small distancing of the category boundaries, and this was not statistically significant, $F(1, 22) = 2.46$, $p > .1$.

Analyses of the questionnaire data showed that 23 of the 24 bilinguals reported English as their dominant language, 91.7% reported talking always or mostly English with friends and acquaintances, and 86.3% reported talking English only or regularly at work. On the other hand, 83.3% reported talking Mandarin only or regularly with relatives. We conclude that our results support the hypothesis that the mechanisms enabling perceptual adaptation to talkers need a regular supply of novel conversational partners; they cannot be maintained in working order if the only available interlocutors are the ones you have been talking to your whole life.

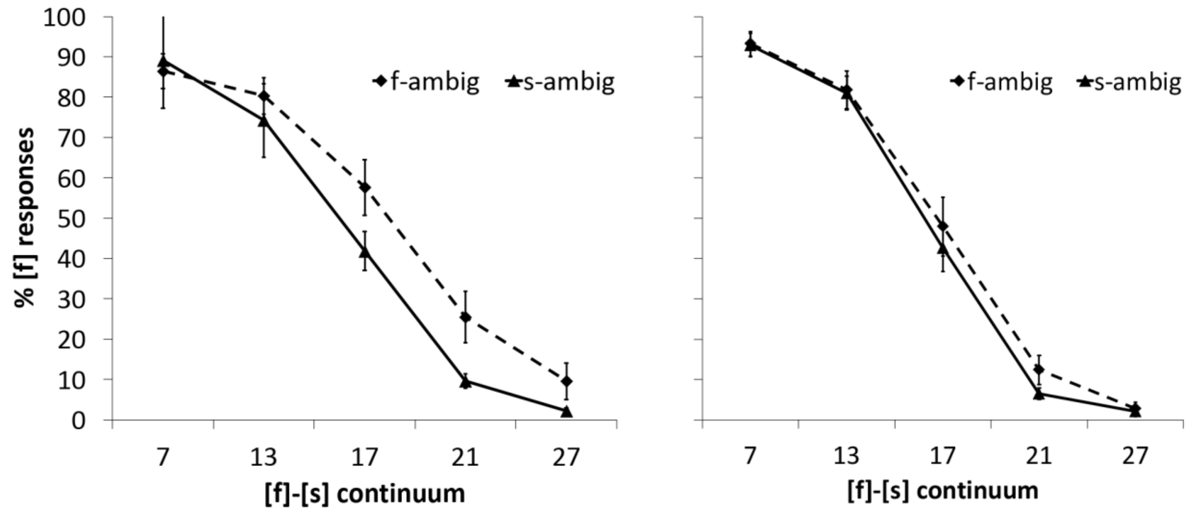


Figure 1. *Phonetic categorization results in English (left) vs Mandarin (right) by the same group of bilingual listeners as a function of training condition. The vertical axis shows % f judgements; the horizontal axis shows the tested 5 points along a 41-step continuum from [f] at the left to [s] at the right. In the English experiment, the training has caused the [f]-[s] category boundaries for the two groups, who heard the ambiguous token replacing [f] vs. [s] respectively, to shift significantly apart from one another; i.e., there has been category boundary readjustment. The difference in the Mandarin experiment is insignificant.*

Acknowlegement

This research was supported by the Australian Research Council (DP 140104389).

References

[1] Norris, D., McQueen, J. M., & Cutler, A. 2003. Perceptual learning in speech. *Cognitive Psychology 47*, 204–238.

[2] Mitterer, H., Chen, Y. & Zhou, X. 2011. Phonological abstraction in processing lexical-tone variation: Evidence from a learning paradigm. *Cognitive Science 35*, 184–197.

[3] Drouin, J. R., Theodore, R. M., & Myers, E. B. 2016. Lexically guided perceptual tuning of internal phonetic category structure. *Journal of the Acoustical Society of America 140*, EL307-EL313.

[4] Reinisch, E., & Mitterer, H. 2016. Exposure modality, input variability and the categories of perceptual recalibration. *Journal of Phonetics 55*, 96-108.

[5] Reinisch, E., Weber, A., & Mitterer, H. 2013. Listeners retune phoneme categories across languages. *Journal of Experimental Psychology: Human Perception and Performance 39*, 75–86.

[6] Schuhmann, K. S. 2014. *Perceptual learning in second language learners*. PhD dissertation, State University of New York at Stony Brook.

[7] Bruggeman, L. 2016. *Nativeness, dominance, and the flexibility of listening to spoken language*. PhD dissertation, Western Sydney University.

[8] Drozdova, P., van Hout, R., & Scharenborg, O. 2016. Lexically-guided perceptual learning in non-native listening. *Bilingualism: Language and Cognition 19*, 914–920.

[9] Cutler, A., Burchfield, L. A., & Antoniou, M. 2018. Factors affecting talker adaptation in a second language. *Proceedings of the Seventeenth Australasian International Conference on Speech Science and Technology* (Coogee, Australia).

Compensation strategies in non-native English and German productions: Evidence for prosodic transfer and adjustment

Jenny Yu[1,2] and Katharina Zahner[3]
[1]The MARCS Institute, Western Sydney University, Australia [2]ARC Centre of Excellence for the Dynamics of Language, Australia [3]University of Konstanz, Germany

German and English speakers differ in how they realise nuclear pitch accents when there is limited sonorant material [1, 2]. English speakers tend to *compress* pitch movements for both falls and rises, i.e., the rate of f0 change increases with decreasing sonorant material. German speakers, however, *compress* rises but *truncate* falls, i.e., rate of f0 change remains stable as sonorant material decreases ([1, 2] for British English vs. Northern German, and Australian English vs. Southern German, respectively). In the present study, we examined whether compensation strategies transfer from native to non-native productions of English and German, i.e., English speakers speaking German and vice versa. This research is important as it may help non-native speakers find more ways to reduce foreign accent and increase intelligibility.

Twelve native Australian English speakers (5 females, M_{age}: 33.9 years, self-rated L2 proficiency (1-7 scale): 4.83) and 12 Southern German speakers (9 females, M_{age}: 25.5 years, self-rated L2 proficiency: 5.27) tested in their L1 in [2] were recorded in their L2 producing questions and declaratives, designed to elicit either a nuclear rising contour on the test word (e.g., Isn't that Mr *Sheafer*? Our new neighbour?), or a nuclear falling contour (e.g., That's Mr *Sheafer*! Our new neighbour!), respectively. Test items were four sets of three equivalent "surnames" in each language, with each surname in a set representing one step on a continuum that varied in scope for voicing (e.g., *Sheafer, Sheaf, Shift* for English, *Schiefer, Schief, Schiff* for German). Following [1, 2], "rate of f0 change" (RoCh), i.e., the f0 excursion (f0 max minus f0 min) in semitones (st) of a fall or rise divided by movement duration, was taken to indicate effects of voicing on f0 movement: Comparing across the continuum (longest to shortest step), a RoCh increase is indicative of compression, while a stable RoCh, of truncation. Compression is further assumed to be accompanied with a stable f0 excursion across steps, while truncation predicts a decrease in excursion from the longest to the shortest step. Figure 1 illustrates the average RoCh (st/ms, left panel) and f0 excursion (st, right panel) for **L2 productions**.

For RoCh, a linear mixed-effects regression model with *language* (English, German, spoken by L2 speakers), *contour* (falling vs. rising), and *step* (1, 2, 3) modelled as fixed factors and *subject* and *word type* as crossed random factors found no interaction between *contour* and *language* ($p=.11$), suggesting rises and falls were not realised differently between languages. **For L2 rises**, RoCh differed between individual steps (all $p<.01$). Specifically, RoCh increased as sonorant material decreased, indicating compression. Analyses of f0 excursion corroborated the use of compression in rises, with a stable f0 excursion for L2 English rises (all individual comparisons $p>.26$) and L2 German rises (all $p>0.28$, except step 1 vs. 3, $p=.03$). **For L2 falls**, RoCh differed between step 2 and 3 ($p<.001$), suggestive of compression between these steps, but, importantly, there was no difference between the shortest and the longest step (1 vs. 3, $p=.09$) nor between step 1 and 2 ($p=.20$), indicative of truncation overall. Truncation was also supported by f0 excursions, which decreased for each step of the continuum (all $ps<0.001$).

Overall, Southern German speakers compressed rises and truncated falls in English (mirroring behaviour in their L1 [2]). Australian English speakers compressed L2 rises as in their L1 in [2], but showed a tendency to truncate falls in German, which differed to the compression patterns found in their L1 in [2]. Therefore, our results paint a complex picture of both transfer and adjustment of compensation strategies from native to non-native productions. To account for this asymmetrical behaviour in prosodic transfer, we will discuss factors known to influence L2 productions, such as markedness (e.g., [3, 4]) and language proficiency (e.g., [5]), as well as a more general factor, such as articulatory effort (e.g., [6, 7]).

Left Panel: Rate of change (RoCh) in L2 | **Right panel: F0 excursion in L2**

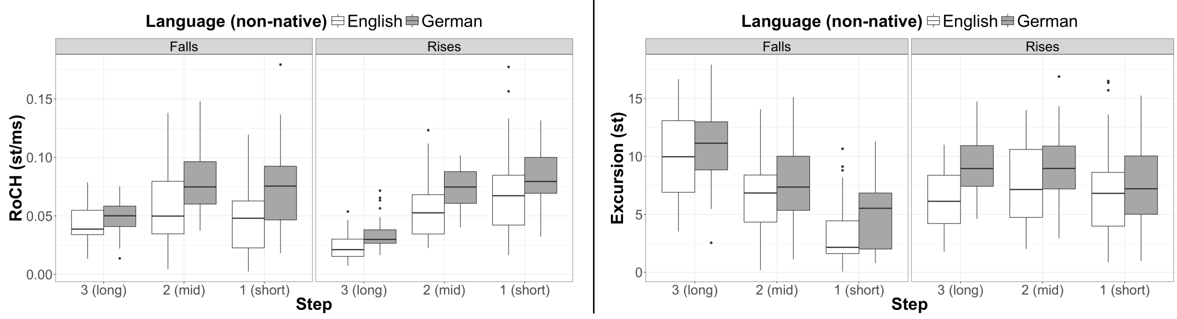


Figure 1. *Rate of Change (RoCh) in st/ms (left panel) and f0 excursion in st (right panel) in different steps for falls (left facet) and rises (right facet), split by language (non-native).*

[1] Grabe, E., "Pitch accent realization in English and German," *Journal of Phonetics,* vol. 26, pp. 129-143, 1998.

[2] Yu, J. and Zahner, K., "Truncation and compression in Southern German and Australian English," in *Proceedings of the 19th Annual Conference of the International Speech Communication Association (Interspeech)*, Hyderabad, India, 2018, pp. 1833-1837.

[3] Zerbian, S., "Markedness considerations in L2 prosodic focus and givenness marking," in *Prosody and Language in Contact: L2 Acquisition, Attrition and Languages in Mulitlingual Situtions*, Delais-Roussarie, E., Avanzi, M., and Herment, S., Eds., Berlin: Springer, 2015, pp. 7-27.

[4] Rasier, L. and Hiligsmann, P., "Prosodic transfer from L1 to L2. Theoretical and methodological issues," *Nouveaux cahiers de linguistique Française,* vol. 28, 2007.

[5] Swerts, M. and Zerbian, S., "Intonational Differences between L1 and L2 English in South Africa," *Phonetica,* vol. 67, pp. 127-146, 2010.

[6] Cohen, A., Collier, R., and Hart, J. t., "Declination: Construct or intrinsic feature of speech pitch?," *Phonetica,* vol. 39, pp. 254-273, 1982.

[7] Gussenhoven, C., *The phonology of tone and intonation*. Cambridge: Cambridge University Press, 2004.

Poster session 1

Default and exceptional stress processing in Spanish: an ERP analysis

Karolina Broś, Martin Meyer and Volker Dellwo

In this paper we set out to investigate the status of stress as an abstract category and its connection with both the phonetic parameters and lexical access. In a neurophysiological experiment measuring event-related potentials (ERPs), we tested the generativist hypothesis concerning lexical storage vis à vis its competing usage-based model [1]. We also tested the hypothesis that in Spanish, in the absence of intonational cues, intensity in the higher regions of the spectrum is of major importance in stress perception [6].

Spanish is characterised by variable stress with an uneven distribution of penultimate, antepenultimate and final-stressed words in the language. Over 64% of all Spanish words are stressed on the penultimate syllable [2] (78.9% in [3]), while antepenults constitute merely 8% (or 2.76%) and should be considered exceptional. This frequency pattern, together with the partial morphological and phonetic predictability of each of the stress types, allows us to assume that there is a default penult pattern derivable by rules with lexical exceptions that have to be learned (see e.g. [4]). At the same time, given variable stress and the existence of minimal pairs, we expect that Spanish speakers are sensitive to stress differences in perception and successfully identify stressed syllables (they are not 'stress-deaf', [5]). Moreover, it was demonstrated that stress perception in Spanish depends on several cues, most importantly pitch, duration and intensity. In non-accented words, spectral tilt was reported to play an important role instead of general intensity differences between stressed and unstressed vowels [6, 7]. With this in mind, we decided to (a) investigate the abstraction of phonetic detail to stress as a phonological category by testing two differently distributed stress patterns, and (b) look at the neurophysiological response to both F0 and spectral tilt changes that affect stress perception. More specifically, we wanted to establish whether the default penultimate stress pattern is processed differently than the exceptional antepenult and whether the latter but not the former is stored in the mental lexicon to facilitate word retrieval (compare other papers studying stress perception with the use of EEG: [8, 9]). To achieve this goal, we must gain access not only to pre-lexical processing, but also to semantic activation responsible for linking phonology with meaning, hence we used a semantic processing EEG paradigm evoking the N400 negativity effect [10]. As a second task, we used a passive oddball paradigm disentangling F0 from intensity cues and designed to evoke a mismatch negativity response (MMN, [11]).

32 native speakers of Spanish (19 females) aged 19-32 listened to 240 stimuli including either correctly or incorrectly stressed trisyllabic words. 60 penults and 60 antepenults with a CV.CV.CV structure were recorded with a female native speaker in two versions and then spliced into an invariable carrier sentence. The words from the two patterns were of matching frequencies and controlled for phonological neighbourhood. Antepenults had a deviant version with penultimate stress and vice versa. A subset of 16 participants took part in the second task, which involved passive listening to a correctly stressed antepenult native word and three randomly distributed deviants, each of which occurred 8% of the time. The first deviant differed from the standard in the F0 of the last syllable, the second differed in the spectral tilt and the third (control) had a vowel change.

The results confirmed the existence of a default stress pattern in Spanish, revealing no N400 effect in the penults as opposed to the antepenults. Accuracy, reaction times and ERPs taken together suggest that stress cues are combined and abstracted to a phonological structure that is either computed online based on grammatical rules (default) or stored in the lexicon together with the representation of the whole word (exceptions). As for the individual cues involved in stress processing in the absence of sentence intonation, the role of spectral tilt was not confirmed. The hearers apparently tune in to F0 to the same extent as to segmental vowel changes, but not intensity shifts.

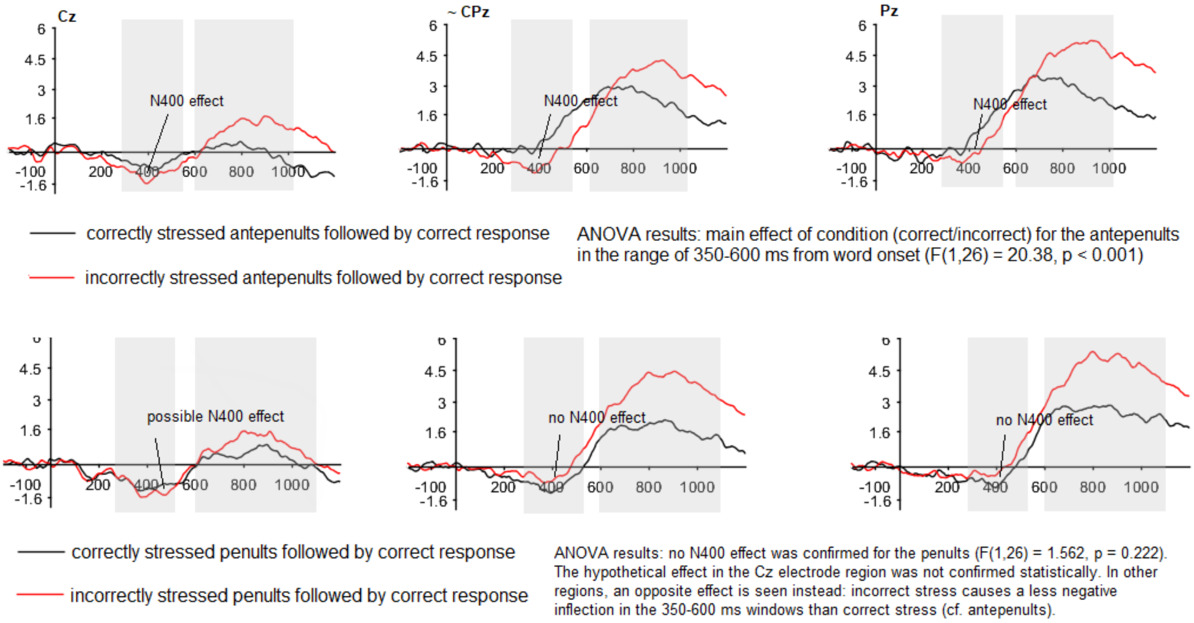


Fig. 1. Summary results of the N400 experiment. Positive values up.

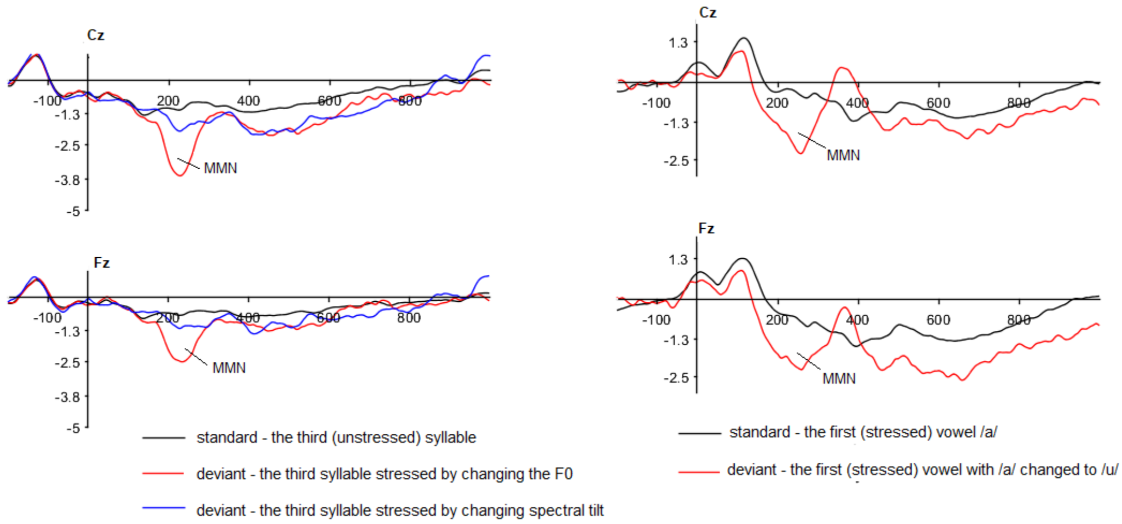


Fig. 2. MMN effect in the case of F0 (left) and vowel changes (right), but not spectral tilt (left, blue line).

References: [1] Bybee, J. (2001). Phonology and language use. [2] Morales-Front, A. (2014). El acento. [3] Quilis, A. (1981). Fonética acústica de la lengua española. [4] Piñeros, C. (2016). The phonological weight of Spanish syllables. [5] Peperkamp, S., Vendelin, I., and Dupoux, E. (2010). Perception of predictable stress: a crosslinguistic investigation. [6] Ortega-Llebaria, M. & Prieto, P. (2007), Disentangling stress from accent in Spanish: Production patterns of the stress contrast in de-accented syllables, [7] Torreira, F., Simonet, M. & Hualde, J. (2014). Quasi-neutralization of stress contrasts in Spanish. [8] Domahs, U., Genc, S., Knaus, J., Wiese, R., Kabak, B. (2012). Processing (un-)predictable word stress: ERP evidence from Turkish. [9] Mołczanow, J. et al. (2013). The lexical representation of word stress in Russian: Evidence from event-related potentials. [10] Kutas, M. and Hillyard, S.A. (1984). Brain potentials during reading reflect word expectancy and semantic association. [11] Näätänen et al. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review.

Word order, focus, and prosodic variation in Nafsan (Vanuatu)

Janet Fletcher, Rosey Billington, Nicholas Thieberger

School of Languages and Linguistics & Centre of Excellence for the Dynamics of Language, University of Melbourne, Australia

In the last ten years or more, there has been an expansion of prosodic analyses of less-well studied languages (e.g. [1, 2, 3]). Compared to well researched European and Asian languages, only a handful of investigations have examined the interaction between prosody and information structure in Oceanic languages with some notable exceptions (e.g. [4] for Samoan). It is generally accepted that tonal variation in languages like English is influenced by a combination of information structure and pragmatics. However, the phonological intonational devices that languages use to contrast informational versus neutral focus are known to vary. These might include combinations of the following: manipulations of phrase-level pitch range (incorporating pitch level and pitch span after [5]), intonational and prosodic phrasing, and intonational prominence, including the use of different types of pitch accents for contrastive emphasis. Languages can also de-accent material (reducing the number of pitch accents in a phrase) and/or de-phrase non-focal material (reducing the number of intonational constituents) to promote a particular kind of discourse interpretation (after [2]).

Languages can use syntactic means to realise informational structure categories like contrastive focus. These devices include left dislocation of the constituent under focus from the rest of an utterance. This has been noted in Oceanic languages (e.g. Vera'a, [6], Nafsan, [7]). It has also been suggested that intonation plays a lesser role in the realisation of semantic focus (compared to west-Germanic languages) with patterns of prosodic variation primarily the result of positional factors. In other words, if a language promotes left dislocation as a contrastive focus-marking strategy, the resulting prosodic patterns may be because the item is in initial position in a discourse segment. By contrast, others have suggested that so-called free-word order languages also employ intonational devices (e.g. [8]) implying that there is a deliberate prosodic strategy to place a constituent in focus. Recent explorations of the complex interplay between prosody, pragmatics, and syntax in Samoan, suggest prosodically-driven syntactic fronting is an important feature of information structure realisation in this language ([4]).

In this paper the interaction between prosody and focus realisation strategies are examined in Nafsan, a Southern Oceanic language spoken by around 5,000 people on the island of Efate in Vanuatu. Nafsan has preferred SVO word order although object fronting is used in cases of topicalisation (after [7]). Our corpus consists of a series of controlled laboratory-phonology type speech experiments that were designed to explore prosodic realisation of neutral and contrastive focus on nouns that were subjects or objects in mini-dialogues where word-order was manipulated. Eight talkers (five males and three females) were recorded in a fieldwork setting. In contexts of contrastive focus, all speakers produce utterance-initial or utterance-final focal elements with a major pitch movement associated with the focused noun (subject or object). The focused noun is also realised with a wider pitch span than the same token in non-focal contexts. They are often realised in their own prosodic phrase and are often prosodically left-dislocated or right-dislocated depending on the position of the noun in the utterance. Post-focal material in Nafsan is almost always produced in a relatively compressed pitch range and there is evidence of de-phrasing of non-focal nouns suggesting prosodic phrasing patterns similar to Korean, for example. Nafsan also exhibits right edge-marking prominence patterns that are amplified in contrastive focus contexts [e.g. 9]. The implication of these findings is considered in relation to prevailing models of prosody and information structure and current models of prosodic typology for languages of the Pacific (e.g. [2,3]).

[1] Steindel-Burdin, R. et al. 2015. Variation in the prosody of focus in head- and head/edge-prominence languages. *Lingua* 165, 254-275.

[2] Jun, S-A. 2014. Prosodic typology by prominence type, word prosody, and macro-rhythm. In S-A Jun (ed.). *Prosodic Typology II*. Oxford: Oxford University Press, 520-539.

[3] Himmelmann, N & Kaufman, D. (in press). Prosodic Systems: Austronesian. In C. Gussenhoven & A. Chen (eds). *The Oxford Handbook of Prosody.* Oxford: Oxford University Press.

[4] Calhoun, S. 2015. The interaction of prosody and syntax in Samoan focus marking. *Lingua* 165, 205-229.

[5] Ladd D.R. 2008. *Intonational Phonology*. Cambridge: Cambridge University Press

[6] Schnell, S. 2018. Whence subject-verb agreement? Investigating the role of topicality, accessibility, and frequency in Vera'a texts. *Linguistics*, 735-780.

[7] Thieberger, N. 2006. *A grammar of South Efate: An Oceanic language of Vanuatu*. Honolulu: University of Hawaii Press.

[8] Zimmerman, M. & E. Onea. 2011. Focus marking and focus interpretation. *Lingua* 121: 1651-1670.

[9] Billington, R., Fletcher, J., & Thieberger N. 2018. Acoustic and durational correlates of prominence in Nafsan. *Proceedings of SST2018*, Sydney December 2018, 4pp.

Using Random Forests to compare production patterns by bilingual English-Spanish speaking children with cochlear implants and their peers with normal hearing

Mark Gibson[1], Ferenc Bunta[2], Charles Johnson[1,3] , Miriam Huárriz[1]
[1]Universidad de Navarra, [2]University of Houston, [3]Brigham Young University

Four random forest models in a supervised learning environment were constructed and used to compare (1) Spanish production data from Spanish- and English-speaking bilingual children with normal hearing (NH group) and Spanish- and English-speaking bilingual children with hearing loss who use cochlear implants (CI group), and (2) the children's productions with the speech of adult native Spanish speakers. The twenty two participants were simultaneous bilinguals from birth. Participants were matched on the basis of chronological age for the NH group (M=5;1) and hearing age for the CI group (M=5;1). The models were fitted for four temporal parameters related to different aspects of syllable structure: (1) lateral duration, (2) vowel duration, (3) voice onset time (VOT), and (4) interplateau intervals (IPI), or the period between two gestural plateaus for a given consonant stop+liquid cluster.

Random forests were preferred over neural networks because they generally require fewer data points to train, and because they have a natural measure for parameter importance based on their use in discriminating between categories. For the current models we used Scikit-learn's ensemble learning package [1] in Python to construct four distinct 100 tree random forest classifiers fitted with articulatory data (electromagnetic articulography, EMA) related to intergestural and oral-laryngeal timing patterns from native Spanish and speaking adults. The trees in each random forest were allowed to have a maximum depth of two and a maximum of six leaf nodes. Two of our classifiers compare adult speech to the CI and NH groups' productions. The other two compare the speech of children with cochlear implants to those without. For each of these two groups, one classifier compares /l/ duration (in /l/ and stop+l contexts), vowel duration and VOT and the other compares interplateau intervals.

Results of the first model show that lateral and vowel duration serve as important parameters when classifying complex onsets, while VOT and IPI are less so (as shown in Table 1), though these latter variables may offer viable cues related to the place categories that can occupy different syllable positions. Additionally, we found that the two classifiers which compare child and adult speech could be thresholded to predict the nature of the speaker with perfect or near-perfect precision for recall and accuracy (see Figure 1). However, the same classifiers performed little better than chance when trying to predict the difference between children with and without cochlear implants (see Figure 2). We interpret the success of these models in the first two instances and their failure in the second two cases to confirm our hypothesis that children with and without cochlear implants in both languages are nearly indistinguishable with respect to the temporal parameters tested.

Results will be discussed in relation to existing work regarding the acquisition of syllable level timing patterns by bilingual children and children with hearing loss who use cochlear implants.

(1) Feature performance values for Spanish timing variables

	NH/CI discrimination				Adult/child discrimination			
	/l/ dur	V dur	VOT	IPI	/l/ dur	V dur	VOT	IPI
C	0.26865	0.54756	0.18377		0.42909	0.54383	0.02707	
CC	0.42860	0.3057	0.11415	0.15153	0.23739	0.70217	0.01682	0.04361

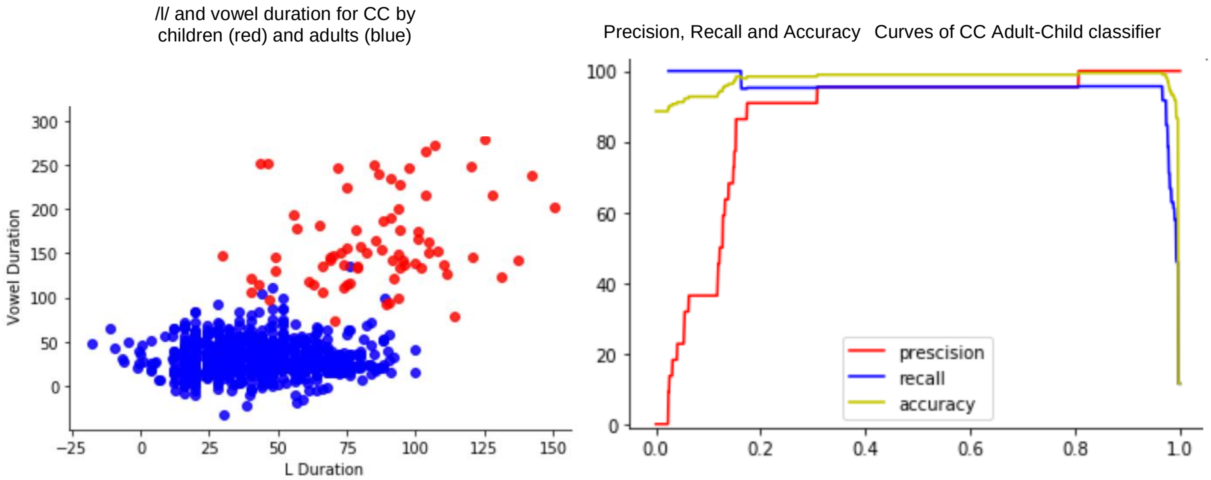


Figure 1. *(left) scatterplot comparing adult (blue dots) and children (red dots) /l/ duration and vowel duration productions in CC clusters. (Right) precision, recall and accuracy curves in discriminating between children and adults based on /l/ duration and vowel duration. The model was able to discriminate children from adults based almost entirely on /l/ duration and vowel duration (though as shown in Table 1, VOT and IPI were also on occasion used).*

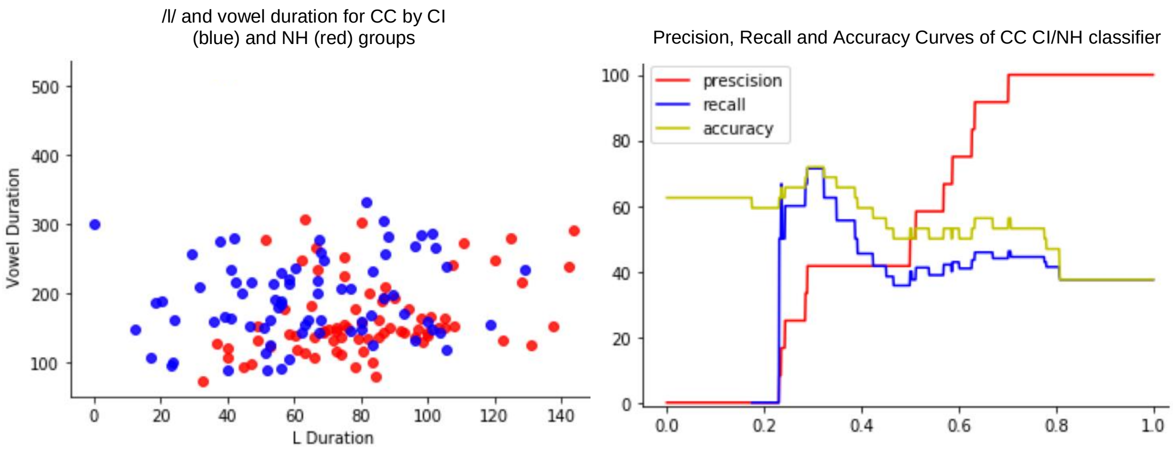


Figure 2. *(left) scatterplot comparing CI (blue dots) and NH (red dots) /l/ duration and vowel duration productions in CC clusters. (Right) precision, recall and accuracy curves in discriminating between CI and NH based on /l/ duration and vowel duration. The model performed no better than chance in discriminating the groups*

[1] Pedregosa et al., 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12, 2825-2830.

Acoustic differences between supposedly identical singular-plural nouns in German

Melina Heinrichs, Marcel Schlechtweg, and Marcel Linnenkohl
University of Kassel, Germany

Words such as *time* and *thyme* and suffixes such as the English plural *-s* and the genitive *-s* have normally been considered homophonous. However, in recent years, several studies have shown that these forms differ in their precise acoustic detail, e.g., in their duration (see, e.g., Drager 2011; Gahl 2008; Plag, Homann, & Kunter 2017). The present contribution aims at expanding research in this area by investigating whether supposedly identical nouns in the singular and plural are really acoustically identical.

Eight native speakers of German (four males) participated in the Praat (Boersma & Weenink 2018) production study presented here. The test material included (a) eight pairs of disyllabic, initially stressed, and monomorphemic German nouns whose nominative singular and plural are expressed with the same word (see 1a and 1b) and (b) eight pairs of disyllabic, initially stressed, and bimorphemic German nouns whose singular and plural genitive are expressed with the same word (see 1c and 1d). The target words of the four conditions were embedded in sentences (see 1), which were read out. Each subject read a total of 32 experimental sentences, i.e., the entire test material (8 items x 2 conditions per item + 8 items x 2 conditions per item), and 64 filler sentences. The four conditions were counterbalanced using a Latin Square Design. Each participant saw a different order of the items. At least 47 other sentences appeared between an item in one condition (e.g., 1a) and the same item in a different condition (e.g., 1b).

All nouns were of a very low frequency. The noun forms of the four conditions were checked for their frequencies, and only relevant cases were manually included in the average frequency. That is, for instance, the form *Spatzen* ‘of the sparrow(s)’ can also occur in cases other than genitive; however, in the average value of the word *Spatzen* in the genitive, only genitive case forms were included. The average frequency values of the four conditions (see, e.g., 1) did not significantly differ from each other. All frequencies were checked using the COSMAS2 (Connexor) corpus of the Institute of the German language (https://www.ids-mannheim.de/cosmas2/).

Using a boxplot analysis (see Larson-Hall 2010), all statistical outliers were removed from the original dataset. Then, repeated-measures ANOVAs by subject and by item were performed on the two dependent variables DURATION INITIAL WORD PART and DURATION FINAL WORD PART. “Initial word part” refers to the part of a word preceding the *en*, i.e., e.g., the stem of the genitive words (see *Spatz* in 1 c/d). “Final word part” refers to the *en*, which was a suffix in (1c/d) but not in (1a/b). The independent/fixed variables were NUMBER (singular/plural) and CASE (nominative/genitive) (both were within-subject, NUMBER was within-item and CASE was between-item). SUBJECT and ITEM were included as random factors. The analysis revealed a significant effect for DURATION INITIAL WORD PART in F_2: Plural nouns were spoken with significantly longer duration than singular nouns ($p < .05$).[1] The results show that plural nouns differ acoustically from their respective singular forms, although the two have been considered to be identical. That is, the “basic” and less marked singular form is uttered in a more compromised form than the plural version. Overall, the results will be interpreted against the background of models of speech production and the role of acoustic detail in the distinction between morphosyntactic properties.

[1] There was no significant interaction and no main effect of CASE.

(1)	a.	*Der*	***Batzen***			*kippt*	*vom*	*Teller.*		
		The	chunk.SING.NOM			falls	from.the	plate.		
	b.	*Die*	***Batzen***			*kippen*	*vom*	*Teller.*		
		The	chunk.PLU.NOM			fall	from.the	plate.		
	c.	*Der*	*Kopf*	*des*	***Spatzen***	*kippt*	*hin*	*und*	*her.*	
		The	head	of.the	sparrow.SING.GEN	goes	back	and	forth.	
	d.	*Der*	*Kopf*	*der*	***Spatzen***	*kippt*	*hin*	*und*	*her.*	
		The	head	of.the	sparrow.PLU.GEN	goes	back	and	forth.	

[1] Boersma, P., & Weenink, D. 2018. *Praat: doing phonetics by computer*. Computer program.

[2] Drager, K. 2011. Sociophonetic variation and the lemma. *Journal of Phonetics* 39(4), 694-707.

[3] Gahl, S. 2008. *Time* and *thyme* are not homophones: The effect of lemma frequencies on word durations in spontaneous speech. *Language* 84(3), 474-496.

[4] Larson-Hall, J. 2010. *A guide to doing statistics in second language research using SPSS.* New York, NY: Routledge.

[5] Plag, I., Homann, J., & Kunter, G. 2017. Homophony and morphology: The acoustics of word-final S in English. *Journal of Linguistics* 53, 181-216.

Articulatory-to-Acoustic relations in Childhood Apraxia of Speech: Vowel production

Giovanna Lenoci, Chiara Celata and Irene Ricci
Scuola Normale Superiore

Speech production in childhood apraxia of speech (CAS) is characterized by impaired motor control phenomena such as errors in the production of consonants, simplifications of clusters and complex syllables, and higher than normal token-to-token variability in the absence of contextual variation ([1]; [2]; [3]). Token-to-token variability is generally measured according to acoustic parameters (e.g. [4]) or in terms of articulatory variation as determined by EMA (e.g. [5]) or UTI data ([6]; [7]). A close understanding of the acoustic and/or articulatory nature of token-to-token variability is crucial not only for rehabilitative purposes, but also for a deeper knowledge of the relation between acoustic and articulatory dimensions of variation (e.g. [8]; [9]).

The current work is an articulatory (UTI) and acoustic study of vowel production by three 10-year-old Italian CAS children and three control peers. We test the following hypotheses: (i) tongue contours from multiple repetitions of the same V target show higher variability in the speech of CAS children compared to control peers (within-category variability); (ii) tongue contours for different vowels are less distinct from each other in CAS speech (between-category distinctiveness); (iii) increased within-category variability and decreased between-category distinctiveness in tongue contours are reflected in the acoustic output (formant frequencies).

A sentence repetition-after-listening task was used to elicit the production of non-words with /a/, /i/ and /u/ as target vowels in CV syllables with /d/ as onset consonant and lexical stress on the target vowel. Each non-word was repeated eight times non-consecutively by each subject. Audio and tongue profiles were recorded via a Mindray UTI system (sampling frequency: 30 Hz) associated to a 65EC10EA microconvex transducer and a Shure unidirectional microphone. The acoustic midpoint of the vowel was chosen for both formant extraction (F1 and F2 automatic extraction through a PRAAT script followed by LPC-based manual correction, then CLIH speaker normalization, [10]) and tongue contour extraction (through semi-automatic tracker implemented in AAA v.2.16.16 by Articulate Instruments Ltd followed by manual correction). Within-category variability was assessed through Nearest Neighbor Distances (NND; [11]): each spline was converted into a set of x,y coordinates and the mean of all the Euclidean distances between each point of one curve and its nearest neighbor on a comparison curve was calculated, through an R script, for each pair of repetitions of each target vowel produced by each speaker. Between-category distinctiveness was assessed through SSANOVAs ([12]) and Bayesian confidence intervals among the three target vowels produced by each speaker. Acoustic variability was assessed by calculating the standard deviation values of non-normalized formants across the eight repetitions of a given vowel for each speaker; acoustic distinctiveness between vowels was assessed for each group (CAS and control children) by calculating the normalized F1 and F2 Euclidean distances obtained for vowel pairs /i-a/, /a-u/ (F1, Euclidean) and /i-u/ (F2, Euclidean) ([13]).

A preliminary analysis shows that hypotheses (i) and (ii) are confirmed, since CAS children show significantly higher within-phoneme variability values and less distinct tongue contours for the three vowel segments compared to control peers. Changes in lingual movements evoke partly similar changes in the acoustic vowel space (see hypothesis (iii)), particularly in the posterior region (/u/ showing a more fronted tongue profile and a higher F2 in CAS speech); Euclidean distances between pairs of vowels were also smaller for CAS children. The results will be discussed with respect to the effects of the pathophysiology on intraspeaker variability and to the articulatory-acoustic relationship.

[1] Marquardt, T. P., Jacks, A., & Davis, B. L. 2004. Token-to-token variability in developmental apraxia of speech: Three longitudinal case studies. *Clinical Linguistics & Phonetics*, 18, 127–144.

[2] Jacks, A., Marquardt, T. P., & Davis, B. 2006. Consonant and syllable structure patterns in childhood apraxia of speech: Developmental change in three children. *Journal of Communication Disorders*, 39(6), 424-441.

[3] Miller, N. 1992. Variability in speech dyspraxia. *Clinical Linguistics & Phonetics*, 6, 77–85.

[4] Sussman, H. M., Byrd, C. T., & Guitar, B. 2010. The integrity of anticipatory coarticulation in fluent and non-fluent tokens of adult who stutter. *Clinical Linguistics & Phonetics*, 25(3), 169–186.

[5] Van Lieshout, P., Davis, B.B., Lipski, M., Namasivayam, A. 2014. The impact of threat and cognitive stress on speech motor control in people who stutter. *Journal of Fluency Disorders*, 40, 93-109.

[6] Lenoci, G. & Ricci, I. 2018. An ultrasound investigation of the speech motor skills of stuttering Italian children. *Clinical Linguistics & Phonetics*, 32(12), 1126-1144.

[7] Frisch, S. A., Maxfield, N., & Belmont, A. 2016. Anticipatory coarticulation and stability of speech in typically fluent speakers and people who stutter. *Clinical Linguistics & Phonetics*, 30, 277–291.

[8] Beckman, M. E., Jung, T. P., Lee, S. H., de Jong, K., Krishnamurthy, A. K., Ahalt, S. C., & Cohen, K. B. 1995. Variability in the production of quantal vowels revisited. *The Journal of the Acoustical Society of America*, 97(1), 471–490.

[9] Iskarous, K., Shadle, C. H., & Proctor, M. I. 2011. Articulatory-acoustic kinematics: The production of American English /s/. *The Journal of the Acoustical Society of America*, 129(2), 944–954.

[10] Nearey, T., 1978. *Phonetic Feature Systems for Vowels*. Bloomington: Indiana University Linguistics Club.

[11] Zharkova, N., & Hewlett, N. 2009. Measuring lingual coarticulation from midsagittal tongue contours: Description and example calculations using English /t/ and /ɑ/. *Journal of Phonetics*, 37, 248–265.

[12] Davidson, L. (2006). Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *Journal of the Acoustical Society of America*, 120(1), 407–415.

[13] Recasens, D., & Espinosa, A (2006). Dispersion and variability of Catalan vowels. *Speech Communication*, 48, 645-666.

Velocity profiles and prosodic variation in younger and older speakers

Jane Mertens[1], Anne Hermes[1] & Doris Mücke[1]
[1]IfL Phonetik, University of Cologne, Germany

Prosodic marking is a complex phenomenon which applies to different levels in the prosodic hierarchy, requiring a fine-tuning of the physical system. This involves modifications of (supra)segmental tiers to make single prosodic constituents such as a syllable or a word stand out [1]. It has been shown that speakers systematically vary between more and less distinct articulations within each utterance, phrase or even within a word [2]. On the textual tier, they mark prosodic structure by producing longer and larger movements of the lips, tongue and jaw in syllables in prominent positions [3-5]. However, as we age, we experience changes at several physiological levels, resulting in inter alia asymmetric velocity profiles of the movements of the limbs and speech organs [6,7]. A common reported effect in studies on aging in speech is a slowing down (segment/word/sentence duration [8-10]). In the present study, we investigated temporal strategies of prosodic marking in the articulatory substance of younger and older speakers. We focus on acoustic syllable durations and the corresponding velocity profiles of vocalic gestures. Therefore, we measured the symmetry of acceleration and deceleration performances during V-to-V articulation in sequences of unstressed-stressed-unstressed syllables. We expect longer durations in stressed compared to unstressed syllables for both age groups. For the older subjects, we expect longer durations as well as asymmetric velocity profiles.

We recorded 7 older native German speakers (2f, 5m; aged 70-80 years) and 7 younger native German speakers (2f, 5m; aged 20-30 years) using electromagnetic articulography (EMA, AG 501). Target words were presented in contrastive focus, structured ˊCV.CV (/bina, pina, dina, tina/), embedded in a carrier sentence (‘Er hat wie**der ‘CV.CV** gesagt’ – *‘He said again.’*). We labelled acoustic syllable boundaries as well as articulatory landmarks for the vocalic gestures in the vertical dimension: onset, peak velocity and target of the accented (V1) and unaccented vowel (V2) as well as the preceding one (V0) in the EMU webAPP [11].

Fig. 1 shows the acoustic durations of the stressed and unstressed syllable. A linear regression analysis (fixed: *age*, *position*; random: *speaker*) revealed an effect of *age* on the acoustic syllable duration ($\chi2(1)=7.2484$; $p=0.007097$) with an interaction of *age* and *position* ($\chi2(1)=56.456$; $p=5.748e-14$) with longer durations for stressed syllables than following unstressed ones and longer durations in older compared to younger subjects. Fig. 2 displays the movement components of the vocalic gestures: (a) acceleration, (b) deceleration and (c) symmetry of these two (values >1 indicate a longer deceleration phase and values <1 indicate a longer acceleration phase). Both groups showed an asymmetrical pattern with rather long deceleration phases into the stressed V1, followed by long acceleration phases into the unstressed V2. The model revealed an effect of *position* ($\chi2(1)=327.63$; $p<2.2e-16$), an interaction of *position* and *age* ($\chi2(1)=122.24$; $p<2.2e-16$), with *age* effects on unstressed V0. For the deceleration phase, there was an effect of *position* ($\chi2(1)=1691.1$; $p<2.2e-16$), an interaction of *age* and *position* ($\chi2(1)=24.676$; $p=4.382e-06$), with *age* effects on stressed V1. The symmetry measure revealed an effect of position ($\chi2(1)=710.35$; $p<2.2e-16$) and an interaction ($\chi2(1)=96.966$; $p<2.23-16$), with *age* effects on stressed V1.

Both age groups highlighted prosodically important information by *longer* acoustic durations in stressed than unstressed syllables, but older speakers show a further slowing down (and also more variability). The velocity profiles were affected by the prosodic pattern in terms of shifting asymmetries in the respective movement components: both groups showed longer deceleration phases than acceleration phases of the vocalic tongue body movements in the stressed syllables, while the opposite was found for unstressed syllables. Segmental effects of the tested /ɐ-i-a/-sequences play an additional role, but cannot explain the shifts ofthe tongue velocities towards and away from the stressed V1. These modifications span the acoustic syllable, mostly affecting the stressed vowel, but also spreading to its left and right, contributing to the sonority expansion of the prominent vowel. This asymmetry was stronger for older than younger speakers indicating an adaption of the (slower) physical control system to the needs of the prosodic system.

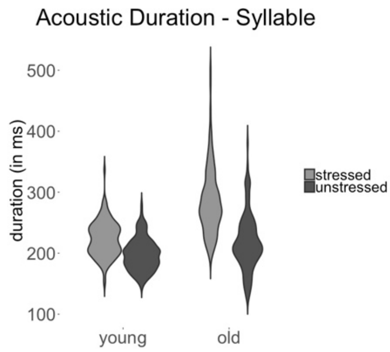


Figure 1. *Acoustic syllable durations (in ms) for younger and older speakers in stressed and unstressed positions.*

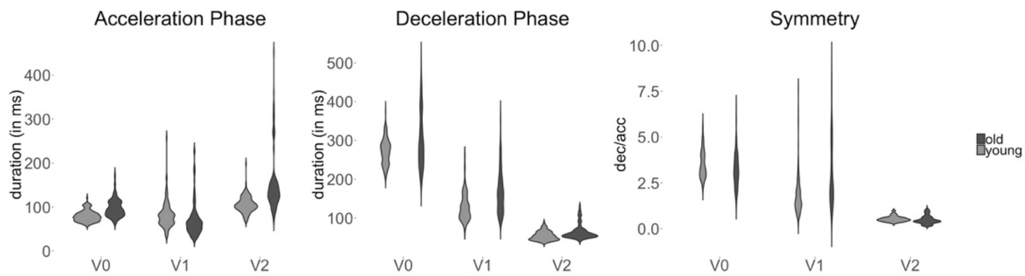


Figure 2. *Duration of (a) acceleration, (b) deceleration phases (in ms) and their (c) symmetry (dec/acc) for younger and older speakers in unstressed V0, stressed V1 and unstressed V2 positions.*

References

[1] Cho, T. (2011). Laboratory Phonology. *The Continuum Companion to Phonology*, 343-368.
[2] Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In *Speech production and speech modelling* (pp. 403-439). Springer, Dordrecht.
[3] De Jong, K. J. (1995). The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *The journal of the acoustical society of America*, *97*(1), 491-504.
[4] Shattuck-Hufnagel, S., & Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of psycholinguistic research*, *25*(2), 193-247.
[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] Ketcham, C. J., & Stelmach, G. E. (2004). Movement control in the older adult. In R.W. Pew, S.B. Vanhemel (Eds.), Technology for Adaptive Aging, National Academies Press, Washington, D.C.
[7] Hermes, A., Mertens, J., & Mücke, D. (2018). Age-related Effects on Sensorimotor Control of Speech Production. *Proc. Interspeech 2018*, 1526-1530.
[8] Amerman, J. D., & Parnell, M. M. 1992. Speech timing strategies in elderly adults. *Journal of Phonetics*, 20(1), 65-76.
[9] Smith, B. L., Wasowicz, J., & Preston, J. 1987. Temporal characteristics of the speech of normal elderly adults. *Journal of Speech, Language, and Hearing Research*, *30*(4), 522-529.
[10] Ramig, L. A. 1983. Effects of physiological aging on speaking and reading rates. *Journal of Communication Disorders*, 16(3), 217-226.
[11] Winkelmann, R., & Raess, G. (2014, May). Introducing a web application for labeling, visualizing speech and correcting derived speech signals. In *LREC* (pp. 4129-4133).

Does the acquisition of reading affect speech production?

A. Popescu[1, 2], L. Hintermeier[1], S. Krüger[1], A. Noiray[1]
[1]Universität Potsdam, [2] Université Paris Diderot – Paris 7

Spoken language acquisition is shaped by multi-faceted developments in the perception, speech motor and phonological domains. While those develop synergistically in most children, certain speech and language disorders have pointed at deviancies in how children organize speech gestures in utterances (*developmental apraxia of speech [6,12]; developmental dyslexia [2,11]; stuttering [4,5]*). In this study, we are interested in the developmental relation that may exist between spoken language and aloud reading. Both mechanisms require a precise control over speech articulators (*e.g. tongue, lips*) as well as their temporal coordination to produce intelligible phonetic outputs. Both mechanisms also require children to develop phonological awareness (PA), the knowledge that words are composed of smaller units (*syllables, segments*). The main difference between the two mechanisms is that spoken language develops much earlier than reading (through repeated exposure and practice speaking the language) while (aloud) reading requires extensive explicit instruction in school.

The goal of the present study is to test whether a link between speech production maturation, phonological awareness and reading exist in children at the early stage of reading acquisition. In particular, we test the hypothesis that PA development and reading observed in primary school lead to significant changes in spoken language production from an organic, experience-based organization to a more precise organization informed by structural knowledge of the native language. The speech production parameter we are considering here is anticipatory V-to-C coarticulation degree (CD), a measure of the temporal overlap of speech gestures that has been found to decrease with age (i.e. from preschool to the first grade to adulthood) with children becoming more proficient speakers [9, 10]. Hence, if our hypothesis is validated, we should observe a negative relation between CD, PA and reading, that is, children with advanced PA and reading skills should exhibit lower CD as a sign of speech pattern maturation resembling those of adults. To test this hypothesis, assessment of CD, PA and reading ability was conducted in 33 native German children at the end of the first grade (mean age 7.04; 20 females), presenting no history of visual hearing or motor disability. The production task consisted in the repetition of pre-recorded disyllabic C1VC2ə non-words by a native female adult model speaker. Target vowels corresponded to a tense vowel (/i:/, /y:/, /e:/, /a:/) and C1 was one of the four /b/, /d/, /g/ and /z/ consonants. Ultrasound imaging was used to record movement of the tongue during the production of the target utterances. Tongue data was analyzed within SOLLAR (a Matlab-based platform developed for kinematic data analysis [8] to extract CD estimates for each child. The raw score of two PA tasks (rhyme production and segment manipulation) as well as reading task (short text) were calculated based on a standard German assessment procedure [3,7]. For reading, a reading fluency score was calculated based on both speed and the accuracy; for PA we selected tasks taping unto the awareness of large (rhymes) and small (segments) units. To test the effect of PA and reading scores of CD, linear mixed models [1] and general additive mixed effects models are used [13,14, 15]. Results show that for both PA and reading scores have non-linear effects on CD, with high scores correlating with lower degree of coarticulation *($p<2e-16$)*. This indicates that the speech of less proficient readers is associated with greater CD and hence organized in syllable-sized units rather than in segments. These findings, should motivate future studies addressing the acquisition of spoken language fluency in languages using alphabetical writing systems to take factors such as PA and reading into account, in addition to other commonly examined factors such as to speech motor control and anatomical development.

References

[1] Bates, D., Mächler, M., Bolker, B., (2015), *Fitting Linear Mixed-Effects Models Using lme4,* Journal of Statistical Software, 67(1): 1-48.

[2] Fawcett, A. J., Nicholson. R.I. (2002), *Children with dyslexia are slow to articulate a single speech gesture.* Dyslexia, 8(4):189-203.

[3] Fricke, S., Schaefer, B. (2008*). Test für phonologische Bewusstheitsfähigkeiten (TPB)*. Idstein: Schulz- Kirchner Verlag.

[4] Hardcastle, W., & Tjaden, K. (2008). *Coarticulation and speech impairment*. In M.J. Ball, M.R. Perkins, N. Müller, & S. Howard (Eds.), The Handbook of Clinical Linguistics (pp. 506–524). Oxford: Blackwell.

[5] Lenoci, G., (2017), *Anticipatory coarticulation in the speech of people who stutter.* On C. Bertini, C. Celata, G. Lenoci, C. Meluzzi & I. Ricci (Eds.), Fattori sociali e biologici nella variazione fonetica – Social and biological factors in speech variation, (pp. 391-409), Milano, Italy: Officinaventuro.

[6] Maas, E., Mailend, ML., 2017, *Fricative contrast and coarticulation in children, with and without speech sound disorders*. Amerivcan Journal of Speech, & Language Pathology, 26(2S): 649-663.

[7] Moll, K., Landerl, K. (2010). *SLRT-II: Lese- und Rechtschreibtest; Weiterentwicklung des Salzburger Lese- und Rechtschreibtests (SLRT)*. Bern: Hans Huber.

[8] Noiray, A., Ries, J., & Tiede, M. (2015). *Sonographic & Optical Linguo-Labial Articulation Recording system (SOLLAR)*, Ultrafest VII, Hong Kong.

[9] Noiray, A., Abakarova, D., Rubertus, E., Krüger, S., & Tiede, M. (2018). *How do children organize their speech in the first years of life? Insight from ultrasound imaging*. Journal of Speech, Language, and Hearing Research, (61): 1355-1368.

[10] Noiray, A., Wieling, M., Abakarova, D., Rubertus, E., & Tiede, M. (2019, forthcoming), *Back from the future: nonlinear anticipation in adults and children's speech,* Journal of Speech, Language and Hearing Research.

[11] Smith, A. B., Smith, S. L., Locke, J. L., & Bennett, J. (2008). *A longitudinal study of speech timing in young children later found to have reading disability*. Journal of Speech, Language, and Hearing Research, (51): 1300-1314.

[12] Terband, H. (2017), *Deviant coarticulation in childhood apraxia of speech (CAS) does not include hyperartion.* Paper presented at the Speech Motor Control conference, Groningen, the Netherlands.

[13] Wieling, M. (2018). *Analyzing dynamic phonetic data using generalized additive mixed modeling: a tutorial focusing on articulatory differences between L1 and L2 speakers of English*. Journal of Phonetics,)70): 86-116.

[14] Wood, S.N. (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

[15] Wood, S.N. (2017) *Generalized Additive Models: An Introduction with R* (2nd edition). Chapman and Hall/CRC.

How prominence shift and phrasing combine to mark focus in French

Cristel Portes[1] and Arndt Riester[2]
[1] *Aix Marseille Univ, CNRS, LPL, Aix-en-Provence, France,* [2] *Universität zu Köln, Institut für deutsche Sprache und Literatur I*

In French, prosodic marking of information structure (IS) is often assumed to be quite different than in flexible accent languages like English or German, where focus is marked by a shift of the main prosodic prominence of the utterance, and givenness is signaled by deaccenting [1]. For instance, [2] proposed that phrasing plays a major role in signaling focus, assuming that French is a boundary language where initial and final rises are boundary tones. In this view, focus constituents are phrased alone and only constituents which are not syntactically embedded, and which are at least the size of the phonological phrase (Φ) can be compressed as a correlate of givenness like in example (8) below where "à Milan" (*at Milan*) can be compressed but not "la bio" (*biology*).

(8) ((Marionω)Φ (enseigneω (la bioω)Φ)Φ (à Milanω)Φ)ι   Φ: phonological phrase
Marion teaches biology in Milan.   ω: prosodic word

This claim is supported by [3] who found no compression on post-focal adjectives in noun+adjectives (N+Adj) segments, and by [4] who found compression on adjuncts but not on arguments in verb+complement segments. Conversely, [5] found that prominence shift is possible in French in "corrective" focus, but not possible in either "contrastive" focus or "parallelism", explaining this finding by the fact that, in French, syntactic scope of the focus operator ~ is minimally the clause.

This study aims at showing how prominence shift is compatible with French phrasing constraints, thanks to the specifics of the intonational structure of French, explaining and discussing both [2, 3, 4] and [5]. It relies on the extensive analysis of a 45 min radio debate, which has been entirely annotated for IS and prosody, independently of one another. For IS, we used an annotation procedure that retrieves the implicit question under discussion (QUD) for each utterance, and defines its focus, focus domain(s), as well as its potential contrastive topic(s), topic(s) and non-at-issue content [6]. For prosody, we used the French ToBI framework [7].

Our data show that both compression on post-focal adjectives in N+Adj segments (against [3], Fig. 1) and prominence shift in contrast and parallelism (against [5], Fig. 2) are regularly used by French speakers to mark givenness and focus respectively. Fig. 2 also shows that focus constituents (here: *toute* and *rien*) are not necessarily phrased alone but can come with post-focal material (against [2]). This is made possible by the intonational structure of French, where the first level of prosodic phrasing, the accentual phrase (AP), is defined as the domain of a compulsory final accent (FA), and an optional initial rise (IR) [8]. The IR has been shown to be often recruited for focusing [9], and the FA, while usually rising, can be realized as a low pitch accent L* [7]. The falling pattern thus formed (Fig. 1 & 2) explains how what [5] have called an "accent shift" is possible in French: enhancing IR and lowering FA, while preserving French prosodic phrasing as required by the phrasing hypothesis [2, 3, 4]. Our data also show how this LHi L* falling pattern is used in building long "accentual arches" [10] that are typically used to signal large focused constituents.

[1] Büring, D. 2016. *Intonation and meaning*. Oxford University Press.

[2] Féry, C. 2001. Focus and phrasing in French. In Féry, C. & Sternefeld, W. (Ed.), *Audiatur vox sapientiae*. Berlin: Akademie-Verlag, 153-181.

[3] Hamlaoui, F., Coridun, S., & Féry, C. 2012. Expression prosodique du focus et du donné au sein des groupes nominaux [NA] du français. In *SHS Web of Conferences* Vol. 1. EDP Sciences, 1505-1518.

[4] Féry, C. 2013. Focus as prosodic alignment. *Natural Language & Linguistic Theory*, *31*(3), 683-734.

[5] Vander Klok, J., Goad, H., & Wagner, M. 2018. Prosodic focus in English vs. French: A scope account. *Glossa: a Journal of General Linguistics*, *3*(1).

[6] Riester, A., Brunetti, L., & De Kuthy, K. 2018. Annotation Guidelines for Questions under Discussion and Information Structure. *Information structure in lesser-described languages.* Amsterdam: John Benjamins, pp. 403-443.

[7] Delais-Roussarie, E. et al. 2015. Intonational phonology of French: Developing a ToBI system for French.

[8] Jun, S. A., & Fougeron, C. 2000. A phonological model of French intonation. In *Intonation.* Dordrecht: Springer, 209-242.

[9] Beyssade, C., Hemforth, B., Marandin, J. M., & Portes, C. 2015. Prosodic Realizations of Information Focus in French. In *Explicit and Implicit Prosody in Sentence Processing*. Heidelberg: Springer Cham, 39-61.

[10] Astésano, C. 2001. *Rythme et accentuation en français. Invariance et variabilité stylistique*. Paris: L'Harmattan.

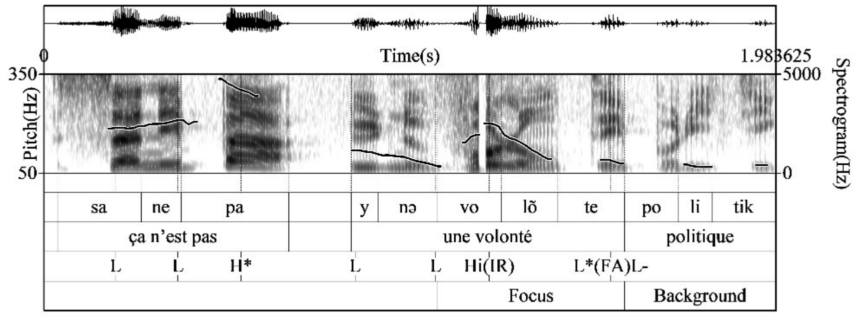


Fig. 1. *The N+Adj segment "volonté politique" (political will) with post-focal compression on the given Adj "politique", realized by means of a low FA L*, followed by the spreading of the phrasal accent L-. Answer to QUD: {Politically, what is the extension of the Union?}*

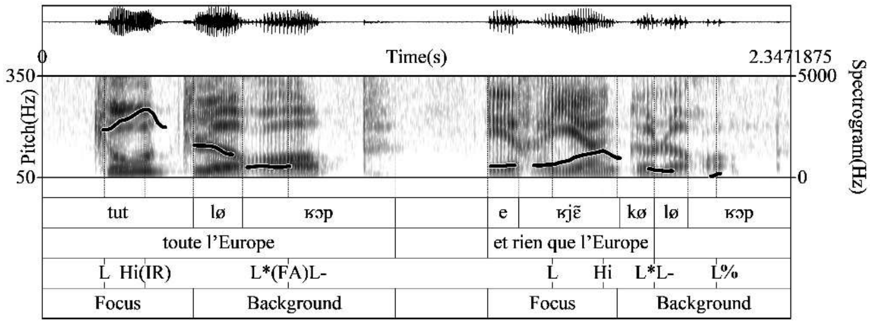


Fig. 2. *The parallelism 'all of Europe, and nothing else than Europe' realized by means of two prominence shifts from the word "Europe" to the words "toute" on the left and "rien que" on the right realized within one AP by means of a low FA L* after a high focused IR Hi. Answer to QUD: {Which parts of Europe should be integrated into the EU?}*

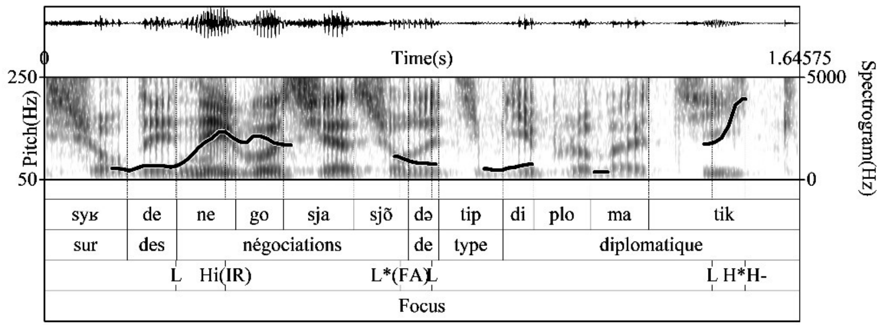


Fig. 3. *The focused DP 'negotiations of a diplomatic kind' realized by means of an "accentual arch". This arch is made of the falling pattern LHiL* on "négociations" plus a rising pattern LLH*H- on "diplomatique". Answer to QUD: {What should we stop founding Europe on?}*

Clusters and complex onsets in *Romagnolo*. A cross-linguistic syllabification algorithm?

Michela Russo Univ. UJM Lyon 3 & UMR 7023 SFL Paris 8/CNRS

In Italian dialects, we have heterogeneous consonant clusters that appear to be branching onsets of (a) increasing sonority /TR/, (b) decreasing /RT/, and (c) equal sonority (TT), what Scheer (2004) defines as a 'plateau' of sonority (see Repetti 1995, Passino 2013, Russo 2016 for the Emilia-Romagna). Particularly, *Romagnolo* is characterised by the deletion of unstressed vowels and the consequent formation of secondary clusters and complex onsets (branching onsets: OL groups Obstruents-Liquids / tl dl zl /.

Analysing the Italian-Swiss maps Atlas from NavigAIS on the Romagna (Ravenna, Forlì-Cesena and Rimini) and Baiolini & Guidetti (2008) for Ferrara (Emilia), we find several secondary branching onsets OL (derived from syncope) with increasing sonority /tl dl zl km tn zn.../ and word-initial complex groups with decreasing RT:

-Branching onsets with increasing sonority – Linear layout TR -

C∅L : ʃã**tl**a lat. médiev. SANC**T(Ŭ)L**US [AIS 36 *madrina/godmother*] - kon**dl**a Vulg. Lat. CŪ**N(Ŭ)L**A [AIS 61 *culla/cradle*] - **zl**owʃ [AIS 66 ***g(e)l**oso/jealous*]- nu**vl**ɑ [AIS 111 *l'u**g(o)l**a/uvula*] - **ʃl**er et **tl**er [AIS 266 *il sellaio/saddler* 1512 *telaio/loom*]

C∅M : **km**er [AIS 36 ***c(o)m**are/godmother*]

C∅N : **tn**aj [AIS 224 ***t(e)n**aglie/pincers*] - **ʃn**urɑ [AIS 49 ***s(i)gn**ora/lady*] - **zn**oʧ [AIS 162 ***g(i)n**occhio/knee*]

Complex groups with decreasing sonority (heterosyllabic ?) - Linear layout RC –

N∅C : (a)**nv**ut AIS 18 *n(i)pote/nephew* - **m**ᵖʃtir AIS 199 *m(e)stiere//craft job*

In the linear layout, I have also included above the branching onsets that are a product of syncope with increasing sonority /tl dl / (like *tler* it. telaio/ 'frame'). These are termed Bogus Clusters by Harris (1994: 182, and ss. for English *choc(o)late, fact(o)ry, myst(e)ry*) and interpreted by Scheer (2004, § 2.6.1), in the framework of the CV strict, as two independent onsets, separated by an empty nucleus.

One argument in favour of a difference between bogus clusters with increasing sonority and branching onsets is distributional. Bogus clusters with a lateral C_2 cannot be found in initial position (*tl, *dl, *gl), for example either in English or Standard Italian. However, this is not the case in *Romagnolo* dialects: type **ʃl**er, **tl**er vs. **tl**ar, **tl**arina it. *telarina*, **tn**aia it. *tenaglia* in Ferrara, etc.. We also find word initial bogus clusters in other Romance languages: in Dolomitic Ladin (*Livinallongo*) and in the dialects of French (see Repetti and Tuttle 1987: 12)
Moreover, (unlike Tuscan Italian), in these dialects lenition applies to those groups identified as bogus clusters /G(U)L/ →/vl/ (nu**vl**ɑ), contrary to what is advocated by the theory GP CVCV, according to which bogus clusters don't undergo lenition, since they always have an empty nucleus to their right. Conversely, in these dialects bogus clusters do lenite. The lenition shows a similarity between the bogus clusters of rising sonority and branching onsets to which phonological lenition also applies in the dialects of Northern Italy. This makes the phonological distinction between bogus clusters and branching onsets non-existent for these dialects.
Furthermore, primary groups with rising sonority can also be split by epenthesis, that is a process of syllabic optimization: in the sequences like /vr dr../ a *svarabhakti* vowel breaks the complex onset /VvRv/.
This phonological process contradicts what is advocated by the theory of the CV strict to the extent that monopositional branching onsets (Lowenstamm 2003; Scheer & Ségéral 2003) may be subject to lenition (Scheer & Brun-Trigaud 2012), but not to epenthesis.

This talk debates the conditions that determine the structuring of clusters and complex onsets:

i. the notion of a syllabic constituent, on which there is no consensus
ii. the principle of sonority: the tendency of segments with more voicing to be heads or ranked by their inclination to dependency
iii. the derived status (non-lexical) of syllabicity

Branching onsets are complex objects that require mechanisms of construction and integration. What is the algorithm that structure phonological branching onsets? Do sonorants have a syllabic role (Sauzet & Brun-Trigaud 2012)?
If one assumes that more voiced segments are intended to be head, one could propose an analysis in which the syllabicity of consonants is a direct possibility and engage in assuming that complex onsets are small CV syllables (Sauzet & Brun-Trigaud 2012), degenerate syllables, not attached to a nucleus (like onsets) : ((CV) (V)). In this sense, a branching onset with increasing sonority reiterates the structure of a syllable. The obstruents + sonorants + /V/ sequence would therefore result from a more abstract pattern: an obstruent + a syllabic sonorant. The dialects of Romagna would then behave like Tashelhit, which does not exclude any segment as the head of a syllable (Dell & Elmedlaoui 1985). This situation does not exclude any sonorant: in the onset /km, tn, mr.../.

This approach develops Kaye & Lowenstamm (1984)'s idea that syllabicity is a status. According to this logic, the syllable is a constituent potentially recursive (Sauzet & Brown-Trigaud 2012; Hulst 2010; 2015) : when segments are small degenerate syllables.

References

AIS = Jaberg, K./Jud, J., 1928–1940. *Sprach- und Sachatlas Italiens und der Südschweiz*, 8 vol., Zofingen, Ringier. NavigAIS = http://www3.pd.istc.cnr.it/navigais-web/

Baiolini, R./Guidetti, F. (2008): *Nuovo vocabolario storico-etimologico del dialetto ferrarese*, Ferrara: Cartografica.

Dell, F. /Elmedlaoui, M. (1985). Syllabic consonant and syllabification in imdlawn tashlhiyt berber. *Journal of African Languages and Linguistics* 7: 105-130.

Hulst, H.G. van der (2010). A note on recursion in phonology. In: Harry van der Hulst (ed.). *Recursion and Human Language*. Berlin: Mouton de Gruyter, 301-342.

Hulst, H.G. van der (2015). But Mr. Lowenstamm, it's CVs all the way down. Conference presented at the congress *The form of structure, the structure of form: three days of linguistics for Jean Lowenstamm à Paris Diderot Paris 7, 15-17 janvier 2015 (ms.).*

Passino, D. (2013). La fonotassi dei dialetti emiliani: osservazioni tipologiche e implicazioni teoriche. Proceeding of the *XVIII Giornata di dialettologia*, Pescarini D. (ed.), juin 2012, Padova, Italie, *Quaderni di lavoro ASIt* n. 16, 57-74.

Repetti, L. and Tuttle, E. F. (1987). The evolution of latin PL, bl, fl, and cl, gl in western Romance. *Studi Mediolatini e volgari* 33: 53-115.

Repetti, L. (1995). Variazione nella sillabificazione: Il caso dei dialetti emiliani e romagnoli' *Rivista Italiana di Dialettologia* (RID) 19: 41-56.

Russo, M. (2016). Les attaques branchantes en anglo-normand et en romagnol : quel algorithme de syllabation pour les attaques complex. Conférence présentée au Workshop *Sonority and Syllabification. Cross-Linguistic evidence and theoretical perspectives*, Lyon (MILC). Journée co-financée par /Workshop co-funded by SFL UMR 7023 CNRS / Université Paris 8 ; LabEx ASLAN /Université de Lyon et ANR Paths (ANR-14-FRAL-0004).

Sauzet, P. & Brun-Trigaud, G. *(2012).* Structure syllabique et évolution phonétique en occitan . In M. Barra-Jover et al. Études de linguistique galloromane. Collections Sciences du Langage. Paris: Presses Universitaires de Vincennes (PUV Université Paris 8), 161-181.

Scheer, T. (2004). *A Lateral Theory of Phnology, vol. 1. What is CVCV and why should it be? Studies in Generative Grammar* [SGG]. Berlin – New York. De Gruyter.

Scheer, T. & Brun-Trigaud, G. (2012). La lenition des attaques branchantes en français et dans les dialects de l'ALF. In Barra-Jover et *al.* (2012), 182-198.

From the eye to the mouth: does a developmental shift in attention co-emerge with the emergence of babbling?

Jennifer Sander[1], Prof. Dr. Barbara Höhle [2],
and Dr. Aude Noiray[14]
[1]*Laboratory for Oral Language Acquisition, Faculty Human Sciences, University of Potsdam, Germany*
[2]*Haskins Laboratories, New Haven, USA*

Speech is a biologically anchored behavior emerging very early in development. However, little is known about the articulatory mechanisms supporting infants' first linguistically-relevant speech exemplars as well as their processing of articulatory information from their environment, in particular in dyadic communicative situations. The present study aims to investigate the interaction of those two processes over developmental time. In the past decades, various studies have outlined the importance of audiovisual integration for speech perception (e.g., [1], [2]), and its production ([3], [4]) in the young age. More recently, a developmental shift in how facial information is processed by infants has been noted, from a focus on the gaze area to increasing attention on speakers' mouth between 8 and 12 months of age ([7]). A relation between both developments has been suggested based on the temporal co-occurrence of babbling and the gaze shift [7].

This study expands on previous findings by investigating German infants' attention to audiovisual (AV) versus unimodal speech (visual only, V) together with their vocal repertoire. We recorded eye movements in 60 monolingual German infants from 6 to 11 months of age in an AV versus V speech perception task. Only babies who did not present any hearing, vision or motor impairment were selected for the study. During the test, infants were sitting on their parents' lap, facing a computer screen on which an eye-tracker (SMI red-m) as well as a video camera were clipped. Infants were presented with audiovisual (AV) or visual (V) exemplars of the isolated vowels /i/ and /a/ pre-recorded by a German native female model speaker. Vowels were chosen because their perception is acquired early in development (i.e., around 6ms, [8]; [9]). The procedure consisted in 5 randomized repetitions of each vowel in each condition. Parents also filled in a detailed questionnaire that included infants' vocal habits. An additional questionnaire assessing infants' vocabulary will be sent at 18 months to test for a developmental relation between attention and lexical growth.

To test for a developmental shift in attention, we determined looks with respect to three regions of interest (ROIs, Figure 1): the eye area, the mouth area, other looks on the face. Using LMERs, we then tested for an effect of experimental condition (AV, V), age (6-7, 8-9, 10-11), vowel (/a/, /i/) and ROI (gaze, mouth, face) on infants´ total looking time (in ms) as the dependent variable. Looking times per ROIs were determined in relation to the total amount of looking time on the face of the model speaker and calculated as a ratio per condition, per vowel, and per child, following the 'eye-mouth index' method used by [10].

Because data analysis is in progress, we only outline our specific predictions. Overall, we anticipate all infants (regardless of age) to look longer at the model speaker´s face (one of the 3 ROIs) in AV condition as compared to V only because this condition approximates natural speech conditions from which infants have been exposed. We expect greater attention to the mouth area in 8-10 months German infants as compared to younger infants (6-7 months of age) and a correlation with their vocal habit as reported in parental questionnaires.

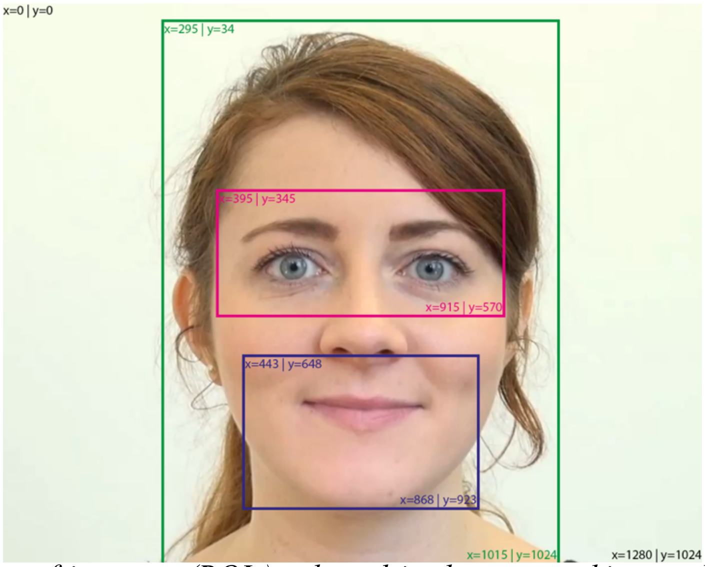


Figure 1. *Regions of interests (ROIs) selected in the eye-tracking study design. Eye area (red), Mouth area (Blue) and additional face area (green).*

REFERENCES

[1] Lalonde, K. & Frush Holt, R. 2015. Preschoolers benefit from visually salient speech cues. *Journal of Speech, Language, and Hearing Research* 58*,* 135-150.

[2] Lewkowicz, D. J. 2000. Infants' perception of the audible, visible, and bimodal attributes of multimodal syllables. *Child development*, *71*(5), 1241-1257.

[3] Kushnerenko, E., Teinonen, T., Volein, A., & Scibra, G. 2008. Electrophysiological evidence of illusory audiovisual speech percept in human infants. *Proceedings of the National Academy of Sciences, 105.* 11442-11445.

[4] Yeung, H. H., & Werker, J. F. 2013. Lip movements affect infants' audiovisual speech perception. *Psychological Science* 24(5*)*, 603-612.

[5] Dodd, B., Mcinosh, B., & Erdener, D. 2008. Perception of the auditory-visual illusion in speech perception by children with phonological disorders. *Clinical linguistics & phonetics* 22(1), 69-82.

[6] Norrix, L. Wl, Plante, E., Vance, R., & Boliek, C. A. 2007. Auditory-visual integration for speech by children with and without specific language impairment. *Journal of Speech, Language, and Hearing Research,* 50, 1639-1651.

[7] Lewkowizc D. J., & Hansen-Tift, A. M. 2012. Infants deploy selective attention to the mouth of a talking face when learning speech. *Proceedings of the National Academy of Sciences* 109(5*), 1431-1436.*

[8] Kuhl, P., Stevens, E., Hayashi, A., Deguchi, T., Kiritani S., & Iverson, P. 2006. Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science* 9(2), F13-F21.

[9] Polka, L. & Werker, J. F.1994. Developmental Changes in Perception of Nonnative Vowel Contrasts. *Journal of Experimental Psychology: Human Perception and Performance,* 20(2), 421-435.

[10] Young, G., Merin, N., Rogers, S., & Ozonoff, Sally. 2009. Gaze behaviour and affect at 6 months: predicting clinical outcomes and language development in typically developing infants and infants at risk for autism. *Developmental Science 12:5, 798-814.*

Quotation marks are expressed in spoken language

Marcel Schlechtweg and Holden Härtl
University of Kassel, Germany

Name-informing constructions (NaInfC) involving predicates like *call* as in *Blood poisoning is also called "sepsis"* are instances of pure quotation, i.e., a meta-linguistic device used to point to linguistic shapes, see, e.g., Cappelen & Lepore (1997). NaInfC inform the addressee about the conventionalized name of a lexical concept, e.g., the name *sepsis*, as opposed to *The doctor diagnosed a sepsis*, where *sepsis* is used denotationally (DenoC). While the semantic and pragmatic properties of (pure) quotation are well explored, its phonetic realization is widely understudied, with only few studies typically examining the acoustic profiles of reported speech, see, e.g., Bertrand, & Espesser (2002); Jansen, Gregory, & Brenier (2001); Klewitz, & Couper-Kuhlen (1999); Oliveira, & Cuhna (2004).

Our paper aims at investigating (i) whether quotation in general is reflected acoustically and (ii) whether the articulator is sensitive to name-informing quotation. For this purpose, we compared the acoustic parameters of NaInfC (see 1a/b) versus DenoC (see 1c/d) and of non-quoted (see 1a/c) versus quoted (see 1b/d) nouns.

In a production study, sixteen native speakers of German were recorded while reading eight German monomorphemic, disyllabic, and initially stressed nouns of low frequency (*Kaper* 'caper', *Pappel* 'poplar', *Kutte* 'robe', *Kippa* 'kippah', *Koppel* 'paddock', *Kuppe* 'tip/peak', *Pita* 'pita', *Pauke* 'timpani') embedded in the four conditions represented in (1). The stressed/initial syllable, i.e., the target syllable, of all nouns was open and composed of a voiceless plosive (= target plosive) as well as a vowel (= target vowel). The sentences of the four conditions were identical, the only differences being the absence/presence of quotation marks and the presence/absence of a name-informing adjective that preceded the noun in focus. The two adjectives, i.e., the name-informing *sogenannte* ('so-called') and the non-name-informing *wohlbekannte* ('well-known') had the same number of morphemes and syllables, and the same stress pattern. Further, the four segments immediately preceding the target syllable of the noun were identical between the two adjectives (see 1). Each subject read all of the thirty-two experimental cases (eight items x four conditions per item) as well as sixty-four filler cases. The order of the four conditions was counterbalanced using a Latin Square Design. Each participant was exposed to a different order. Twenty-three other sentences appeared between an item in one condition and the same item in another condition.

The data was analyzed with Praat and several dependent variables were considered: (A) duration of target syllable, (B) duration of target plosive, (C) duration of constriction of this plosive, (D) VOT of this plosive, (E) duration of target vowel, (F) maximum intensity of this vowel, and (G) maximum F0 of this vowel. Repeated-measures ANOVAs by subject and by item were performed, including the fixed factors QUOTATION MARKS (yes/no) and NAME-INFORMING ADJECTIVE (yes/no), their interaction as well as the random factors SUBJECT and ITEM. The two fixed/independent variables were within-subject/item factors.

Quoted nouns (see 1b/d) were pronounced with significantly longer (A), (B), (C), (D) and a significantly higher (G) than non-quoted nouns (see 1a/c). In addition, NaInfC (see 1a/b) showed a significantly longer (B) than DenoC (see 1c/d).[1] Based on the two main effects, we argue that NaInfC are articulated differently than DenoC and, crucially, that the presence/absence of quotation marks has an influence on the acoustic realization of an item. To conclude, we will discuss the implications of our results for theories of quotation as well as the interface between semantics, pragmatics, and phonetics.

[1] There was no significant interaction.

(1) a. *Viele Mönche tragen die* ***sogenannte Kutte*** *täglich von morgens bis abends.*
‘Many monks wear the **so-called robe** everyday from morning to night.’

b. *Viele Mönche tragen die* ***sogenannte „Kutte“*** *täglich von morgens bis abends.*
‘Many monks wear the **so-called “robe”** everyday from morning to night.’

c. *Viele Mönche tragen die* ***wohlbekannte Kutte*** *täglich von morgens bis abends.*
‘Many monks wear the **well-known robe** everyday from morning to night.’

d. *Viele Mönche tragen die* ***wohlbekannte „Kutte“*** *täglich von morgens bis abends.*
‘Many monks wear the **well-known “robe”** everyday from morning to night.’

[1] Bertrand, R., & Espesser, R. 2002. Voice diversity in conversation: A case study. In Bel, B., & Marlien, I. (Eds.), *Speech Prosody 2002*. Aix-en-Provence, France, 171–174.

[2] Cappelen, H., & Lepore, E. 1997. Varieties of Quotation. *Mind* 106(423), 429–450.

[3] Jansen, W., Gregory, M. L., & Brenier, J. M. 2001. Prosodic correlates of directly reported speech: Evidence from conversational speech. *Prosody in Speech Recognition and Understanding* (Paper 14). Red Bank, NJ.

[4] Klewitz, G., & Couper-Kuhlen, E. 1999. *QUOTE – UNQUOTE? The role of prosody in the contextualization of reported speech sequences* (= Interaction and Linguistic Structures 12). Konstanz: University of Konstanz.

[5] Oliveira, M. Jr., & Cunha, D. A. C. 2004. Prosody as marker of direct reported speech boundary. In Bel, B., & Marlien, I. (Eds.), *Speech Prosody 2004*. Nara, Japan, 263–266.

Allophony and history of Mee velar lateral

Peter Staroverov Sören E. Tebay
Wayne State University Leipzig University

Velar laterals are a rare class of sounds that involve posterior closure and lateral release (François, 2010; Ladefoged et al., 1977; Blevins, 1994). This paper presents new data on velar lateral allophony in the Paniai Lakes Papuan language Mee (a.k.a. Ekari: Doble, 1987; Hyman & Kobepa, 2013, a.o.). The velar lateral sound appears with dorsal closure and lateral release [g$^{\text{L}}$] before front vowels and with uvular closure and fricative release [ɢ$^{\text{ʁ}}$] before back vowels. This allophony pattern is documented here for the first time, and it has interesting implications for the historical sources of velar laterals. [ɢ$^{\text{ʁ}}$] is also reported as a marginal sound in Xumi (Chirkova & Chen, 2013).

Data. Our main dataset comes from a male speaker of Mee. The patterns of velar lateral allophony were verified in elicitations with one additional speaker. Mee consonant inventory includes /p b t d k g m n w j/. The velar lateral corresponds to /g/ in this system (see also Doble, 1987), and therefore we propose to treat it as an affricate rather than pre-stopped lateral or a sequence of two sounds (cf. François, 2010). Mee has five vowels /i e a o u/, each contrasting in length, and five diphthongs: /ei ai eu au ou/. Mee only allows CV syllables. The tonal contrasts are analyzed by Hyman & Kobepa (2013).

Examples in ((1), tones omitted) show that /g/ is pronounced [g$^{\text{L}}$] before front vowels and diphthongs /ei eu/ (1a) and as [ɢ$^{\text{ʁ}}$] before back vowels and /ai au ou/ (1b). Short /i e/ are reduced and highly lateralized after [g$^{\text{L}}$] (hence a breve sign in (1c)).

(1) a. [g$^{\text{L}}$eːg$^{\text{L}}$eː] 'to dry in the sun'; [jug$^{\text{L}}$ei] 'to crush'; [jag$^{\text{L}}$iː] 'to fall'
b. [ɢ$^{\text{ʁ}}$aːti] 'ten'; [daɢ$^{\text{ʁ}}$u] 'room'; [eɢ$^{\text{ʁ}}$ou] 'to pull'
c. [g$^{\text{L}}$ĭdiː] 'to take out' [g$^{\text{L}}$ĕmoː] 'cool'; [ag$^{\text{L}}$ĕ] 'floor, ground'; [dag$^{\text{L}}$ĭ] 'head'

We recorded a set of words with /g/ in each vocalic environment, each repeated three times in a carrier phrase [itoko __ natidodou] 'say __ now'.

Results. The release of /g/ was found to be acoustically different before front vs. back vowels. Compared to the sonorous release for [g$^{\text{L}}$], [ɢ$^{\text{ʁ}}$] release is less periodic since it is fricated and sometimes partially devoiced. To assess release periodicity, harmonics to noise (HtoN) ratio within 20ms of the closure offset was measured. [g$^{\text{L}}$] showed more periodic release (mean HtoN 7.08 dB for [g$^{\text{L}}$] vs. 5.47 dB for [ɢ$^{\text{ʁ}}$]), and this difference was significant based on lme model with item as random effect ($\beta = 1.75; SE = 0.45; p < 0.001$)

To assess a difference in /g/ constriction location, formant transitions into /g/ were measured at the 9/10 of V1 duration. The resulting vowel space is in Figure 1. F2 transitions into /g/ were analyzed with a lme model taking V1 quality and V2 frontness as fixed effects and item as a random effect. The significant V2 frontness effect ($\beta = 210; SE = 67; p < 0.01$) is compatible with [ɢ$^{\text{ʁ}}$] closure being further back than that of [g$^{\text{L}}$]. An expected significant effect of V1 quality was also found for all vowels. The interaction between V1 being [e] and V2 frontness was also significant. Overall, the results confirm our description of /g/ allophony: [ɢ$^{\text{ʁ}}$] and [g$^{\text{L}}$] differ in both closure and release.

Discussion. While across languages velar laterals may emerge from stops (Tebay, 2018) or from rhotics (François, 2011), the emergence of lateral release quality is a puzzle. The connection between release quality and vowel frontness in Mee allows us to hypothesize that consonant-vowel coarticulation may be the source of modified release quality in these sounds. We suggest that a stage where consonant release quality changes with the following vowels (as in Mee) could be a precursor to the emergence of a general phoneme realized as velar lateral across the board (as e.g. in Hiw, François (2010)). This hypothesis remains to be tested with additional phonetic data on velar laterals.

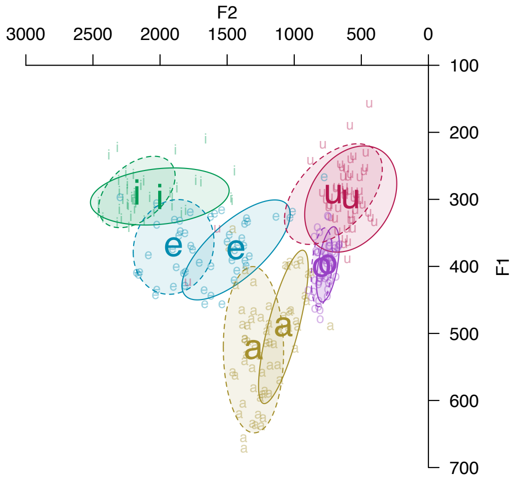


Figure 1: Formant transitions from V1 into [gᴸ] (dotted line) and [ɢʁ] (solid line).

References

Blevins, Juliette. 1994. A place for lateral in the feature geometry. *Journal of Linguistics* 30. 301 – 348.

Chirkova, Katia & Yiya Chen. 2013. Xumi (part 1): Lower Xumi, the variety of the lower Shuiluo river. *JIPA* 43(3).

Doble, Marion. 1987. A description of some features of Ekari language structure. *Oceanic Linguistics* 26(1-2). 55–113.

François, Alexandre. 2010. Phonotactics and the prestopped velar lateral of Hiw: resolving the ambiguity of a complex segment. *Phonology* 27. 393–434.

François, Alexandre. 2011. Where *R they all? The geography and history of *R-loss in Southern Oceanic languages. *Oceanic Linguistics* 50(1). 140–197.

Hyman, Larry & Niko Kobepa. 2013. On the analysis of tone in Mee (Ekari, Ekagi, Kapauku). *Oceanic Linguistics* 52(2). 307–317.

Ladefoged, Peter, Ann Cochran & Sandra Disner. 1977. Laterals and trills. *JIPA* 7(2). 46–54.

Tebay, Sören. 2018. Towards a reconstruction of Proto-Paniai Lakes. Ms., Leipzig Universtiy.

The L1 of Polish emigrants in Denmark – a comparison of three emigration waves

Jolanta Sypiańska
Koszalin University of Technology, Poland

The study of how and why the native language of a healthy adult may differ from a recognised and accepted standard has been approached from a number of perspectives including language attrition (Schmid 2004, Schmid and Köpke 2008), multi-competence (Cook 2003) or cross-linguistic influence in the form of reverse transfer from the foreign language(s) (e.g. Chang 2010, 2012). In the present study, the latter approach is adopted. The particular aim is to decide whether regressive transfer from L2 Danish and L3 English on L1 Polish in three waves of emigrants is conditioned by each wave's different attitude towards the native language and home country. The first wave included Poles of Jewish origin who were deported or otherwise forced to leave Poland following the 1968 March events and the anti-Jewish smear campaign of the late 1960s and early 1970s (n=12). The second wave of emigration comprised refugees who fled the country in the 1980s with the outbreak of political repression against Solidarność activists (n=10). The last wave can be subsumed under the category of labour migrants who came to Denmark of their own accord in the search of better paid jobs and more comfortable lives in the 1990s and early 2000s (n=8).

Attitude has been claimed to be a key factor influencing the amount of L1 attrition (Köpke and Schmid 2004; Schmid 2011) however studies based on questionnaires (e.g. Attitude and Motivation Test Battery used by Schmid and Dusseldorp (2010)) tend not to find significant links between attitude towards the L1 and amount of influence on the L1. That is why in the present study an interview was run to tap the attitudes of the study subjects towards their home country and native language in qualitative terms. The Poles deported due to their Jewish origin had a clearly affronted attitude towards Poland. They were hurt to be asked to leave the country they grew up in. The second wave included people who actively stood up against the Communist regime in Poland. Their attitude towards their homeland was much more favourable as they blamed the communists and not Poland as such for their need to flee the country. The third wave had the most positive attitude towards Poland and treated their stay in Denmark only as a means of bettering their financial situation. Interestingly, the first wave showed a very positive attitude to the native language and the second and third wave a neutral attitude. The differences in the attitudes of the emigrant waves were hypothesised to condition the magnitude of foreign language influence (L2 Danish in the bilingual group and L2 Danish and L3 English in the multilingual group) on their L1 Polish with the first wave showing the most and the last wave the least influence. The influence of the foreign languages was determined by means of comparing the first and the second formant (F1, F2) of the Polish /ɛ/ with a control group consisting of Polish natives residing in Poland. It was visible in the form of a higher and fronter Polish vowel that was clearly becoming more similar to the closest Danish vowel set /e,ɛ,æ/ in the Bilingual group and Danish /e,ɛ,æ/ and English /ɛ/ in the Multilingual group.

As a first step in the analysis, no main effect of length of residence, age of onset or L2 and L3 proficiency were found in the data. The factor of emigrant wave was analysed by means of a one-way ANOVA with emigrant wave (I,II,III) as independent variable and F1 and F2 measured at vowel midpoint as dependent variables. Two-syllable words with initial stress were extracted from spontaneous conversations with the participants and the formants were measured in PRAAT (Boersma and Weenink 2019). The results give evidence of the influence of emigrant wave on F2 ($F=6{,}73613$, $p<.001$) but not on F1 ($F=2{,}23717$, $p=.11$). The observed tendency was for F2 to rise from the first to the third emigrant wave making it more similar to the Danish sounds but with no statistically significant difference between the second and the third wave. The conclusion is that a positive attitude towards the native language may trump a negative attitude towards the home country in conditioning the amount of regressive transfer to the L1. On the other hand, a positive attitude towards the L1 country together with a neutral attitude towards the L1 are not able to block it.

[1] Boersma, P. and Weenink, D. (2019). Praat: doing phonetics by computer [Computer program]. Version 6.0.49, retrieved 2 March 2019 from http://www.praat.org/

[2] Chang, C.B. 2010. *First Language Phonetic Drift During Second Language Acquisition*. [Unpublished PhD at the University of California, Berkeley].

[3] Chang, C.B. 2012. Rapid and multifaceted effects of second language learning on first language speech production. *Journal of Phonetics* 40: 249-268.

[4] Cook, V.J. (ed.). 2003. *Effects of the second language on the first*. Clevedon: Cromwell Press.

[5] Köpke, B. and Schmid, M.S. (2004). Language attrition: The next phase. In M.S. Schmid, B. Köpke, M. Keijzer & L. Weilemar (eds.), First language attrition: Interdisciplinary perspectives on methodological issues (pp. 1-43). The Netherlands: John Benjamins.

[6] Schmid, M. 2004. First language attrition: The methodology revised. *International Journal of Bilingualism* 8, 3: 239-255.

[7] Schmid, M.S. (2011). Language attrition. New York: Cambridge University Press.

[8] Schmid, M.S. and Dusseldorp, E. (2010). Quantitive analyses in a multivariate study of language attrition: The impact of extralinguistic factors. *Second Language Research, 26*(1), 125-160.

[9] Schmid, M., & Köpke, B. 2008. L1 attrition and the mental lexicon. In Pavlenko, A. (Ed.), *The bilingual mental lexicon: interdisciplinary approaches*. Clevedon: Multilingual Matters, 209-238.

Oral session 2

From short term accommodation to change II

Short-term accommodation of Hong Kong English towards RP and GenAmE

Wenling Cao (Grace) University of York, UK

The interactive-alignment model (Pickering & Garrod 2004) suggests that convergence is an automatic process in a conversation and it occurs not only at phonetic level but also at syntactic and pragmatic levels. Evidence of convergence at phonetic level was found in a few studies (Babel 2010, 2012; Pardo 2006; Pardo, Gibbons, Suppes & Krauss 2012), however, these studies mainly focused on native speakers, convergence between non-native speakers and native speakers received less attention. Present study aims to examine Hong Kong English (HKE) speakers' speech accommodation towards Received Pronunciation (RP) and General American English (GenAmE) during and after 1 hour's conversation with an RP/GenAmE interlocutor.

Nineteen HKE speakers conducted a Map Task with a native speaker of RP and a native speaker of GenAmE respectively. A pre-task and a post-task were also conducted to capture the changes in pronunciation. Two vowels (BATH and THOUGHT vowels) and three consonants (fricatives /z/, /θ/ and rhoticity) were chosen as target sounds. F1 and F2 values at the midpoint of the vowels in the pre-task, map task and post-task were extracted for the calculation of means and Euclidean distance. Percentages of the participants' realisation of [z], [θ] and rhoticity in the three tasks were also calculated.

According to the interactive-alignment model (Pickering & Garrod 2004), convergence between the HKE speakers and the native speakers were expected: (1) for the **vowels**, the HKE participants were expected to converge their vowels to be more British-like in the RP condition; their vowels were expected to be more American-like in the GenAmE condition; (2) for **rhoticity**, they were expected to produce less rhotic words in the RP condition and to produce more rhotic words in the GenAmE condition; (3) for the **fricatives**, they were expected to produce less HKE variants of the fricatives (e.g. HKE variant [s] for fricative /z/ and HKE variant [f] for fricative /θ/) and pronounce more [z] and [θ] instead.

Results supported some of the predictions. The HKE participants converged towards the native accents on some sounds but not on the others. Significant convergence from the pre-task to the map tasks was found on fricative /z/ and rhoticity. The participants produced more fricative /z/ when they talked to the native speakers and changed the percentages of rhoticity depending on the accents they were exposed to in the map tasks. However, the HKE participants diverged on fricative /θ/ from the pre-task to the map tasks and no convergence was found on the two vowels. Significant divergence was found on the BATH vowel and no change was found on the THOUGHT vowel. These results only partially support the automatic account claimed by the interactive-alignment model (Pickering & Garrod 2004).

The selectivity of phonetic convergence observed in the present study was also found in Babel (2010, 2012). The reason might be that for the HKE participants some target sounds were more salient than others due to the influence of their L1-Cantonese. For example, phonetic differences between the native variant and the HKE variant for fricative /z/ (i.e. [z] vs [s]) are larger than the differences between those for fricative /θ/ (i.e. [θ] vs [f]). On the other hand, sounds with higher frequency were more likely to accommodate too. Analysis suggested that the HKE participants in average received more native input on the two target sounds which showed convergence (i.e. rhoticity and fricative /z/).

Another interesting finding of the study was that talker gender/sex did not seem to affect people's convergence. The female and male participants had no differences on their convergence on all the target sounds except for fricative /θ/. This result was in line with Pardo, Urmanche, Wilman and Wiener (2017) and challenges the traditional view of accommodation that females accommodate more than males (Namy, Nygaard & Sauerteig 2002).

References:

Babel, M. (2010). Dialect divergence and convergence in New Zealand. *Language in Society*, *39*(4), 437–456.

Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics, 40*(1), 177–189.

Namy, L. L., Nygaard, L. C., & Sauerteig, D. (2002). Gender differences in vocal accommodation: The role of perception. *Journal of Language and Social Psychology*, 21(4), 422–432.

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

Pardo, J. S., Gibbons, R., Suppes, A., Krauss, R. M. (2012). Phonetic convergence in college roommates. *Journal of Phonetics, 40*(1), 190–197.

Pardo, J., Urmanche, A., Wilman, S., & Wiener, J. (2017). Phonetic convergence across multiple measures and model talkers. *Attention, Perception, & Psychophysics*, 79(2), 637–659.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *The Behavioral and Brain Sciences*, 27(2), 169-190.

Phonetic accommodation on the segmental and the suprasegmental level of speech in native-non-native collaborative tasks

Christiane Ulbrich
University of Konstanz

Numerous studies have previously shown interaction of both phonetic and phonological categories of two (or more) languages in the speech obtained from multilingual individuals. Evidence for a phonetic drift and/or accommodation effects of systematic characteristics has frequently been presented in recent publications (Flege, 1995; Sancier & Fowler, 1997; Jiang, 2010; Fabiano-Smith, Viswanathan, Olmstead, 2010). Observed changes are usually attributed to aspects of bi-directional influences between languages. However, they fail to account for inconsistent accommodation effects. In addition, issues still debated in this context are (1) the question whether or not experience and language mastery have an effect on the extent of such changes (Chang, 2013), (2) what source causes these phonetics drifts or changes (Tobin, 2015) and a (3) what levels of spoken language representations may be affected most, segmental characteristics or suprasegmental characteristics.

This paper presents an investigation of productions data obtained from 12 native Spanish speakers of L2 German and 6 native German controls. Recordings were done during a collaborative map-task performed by a Spanish or by a mixed pair of interlocutors. The Spanish-Spanish map-tasks were carried out twice, in one map-task we recorded speakers of the same proficiency level and in a second we recorded the interaction between one highly proficient speaker and a beginner. The map task was designed in a way that specific target words and phrases appeared in the beginning and at the end of the collaborative tasks. We analysed several aspects of the speech signal, derived from contrastive analyses of Spanish and German (Lleó et al., 2003; Hirschfeldt, 1988):

(i) /r/-realisation
(ii) neutralisation of final voicing contrast in plosives
(iii) realisation of initial /h/
(iv) intensity, duration and pitch range and pattern of accented vs. unaccented syllables
(v) speech and articulation rate

Two main objectives were pursued in the study: (1) we aimed to shed light on the question whether there is a decreasing interference of Spanish phonology on L2 German productions in the utterances produced by speakers of higher proficiency levels and (2) whether segmental and suprasegmental characteristics are affected to different degrees. To allow for a quantitative assessment of short-term drifts speech samples were segmented, annotated and acoustically analysed using PRAAT. The statistic analyses shows that both L1 and proficiency level of the interlocutor have an effect. However, we found that not all aspects are affected by the accommodation effect to the same extent. In most highly proficient speakers' utterances a shift from tapped and/or trilled /r/ to more fricative like /r/ realisations was found. Neutralisation of the final voicing contrast and was not found whereas some speakers started to produce the glottal fricative in initial position however with more spectral energy compared to the German target. The perceptual relevance was tested in an XAB similarity judgment task carried out by 17 native speakers of German. X was a selected utterance (word or phrase) from one speaker. A and B were early and late utterances from the interlocutor. Phonetic drift was observed in native-native and non-native-native conversations. However, in conversations between less proficient non-natives with native Speakers of German drift was indexed to a lesser extent and none of the native speakers converged towards a non-native partner, Greatest convergence was found in conversations of highly proficient non-natives with native speakers of German. The results suggest that phonetic accommodation can occur cross-linguistically, and that it may be constrained both by speakers' dialect and by their language proficiency. In line with previous findings the results can best be account for by an adaptation of a dynamic system approach.

References

Fabiano-Smith, L., Viswanathan, N., Olmstead, A. 2010. Short term phonetic and phonological changes during Spanish instruction.. J. Acoust. Soc. Am. **128**, 2486.

Hirschfeld, U. 1988. Sprachvergleich Spanisch-Deutsch im Bereich der Phonetik/Phonologie. In: Wotjak, Gerd; Regales, Antonio (eds.). Studien zum Sprachvergleich Deutsch-Spanisch. Valladolid : Spanischer Deutschlehrerverband, pp. 31-36.

Jiang, H., 2010. Effect of L2 phonetic learning on the production of L1 vowels: A study of Mandarin-English bilinguals in Canada. In K. Dziubalska-Ko laczyk, M. Wrembel, and M. Kul (Eds.), New Sounds 2010: Proceedings of the 6th International Symposium on the Acquisition of Second Language Speech, Poznan, Poland, ´ pp. 227–232. Adam Mickiewicz University.

Lleó, C., Kuchenbrandt, I.; Kehoe, M., Trujillo, C., 2003. Syllable final consonants in Spanish and German monolingual and bilingual acqusition. In. Müller, Natascha (ed.). (In)vulnerable domains in multilingualism. Amsterdam: Benjamins, pp. 191-220.

Tobin, S., 2015. A dynamic approach to phonetic change. In *Proceedings of the International Congress of Phonetic Sciences* (M. Wolters, J. Livingstone, B. Beattie, R. Smith, M. MacMahon, J. Stuart-Smith & J. Scobbie, Eds.). London: IPA.

Chang, C. B., 2013. A novelty effect in phonetic drift of the native language Journal of Phonetics, 41(6) 520–533

Sancier, M. L., & Fowler, C. A., 1997. Gestural drift in a bilingual speaker of Brazilian Portuguese and English. Journal of Phonetics, 25(4), 421-436.

Poster session 2

Are people as old as they sound? Acoustic, regional and generational effects

Nicolas Audibert and Cécile Fougeron
Laboratoire de Phonétique et Phonologie, UMR7018 CNRS/Sorbonne-Nouvelle, Paris

The age of a speaker is part of the various social information which can be indexically signaled in his/her speech (Foulkes & Docherty 2006; Foulkes et al. 2010; Ekert 1997). A recurrent question is whether age should be defined in chronological, social, cognitive or biological terms. Here, we address the question of the relationship between 'chronological' and 'perceived' age, and the different factors that could explain how and why they differ.

Chronological and perceived age are compared in the speech of 112 speakers of French aged from 50 to 89 y.o., distributed into 4 regional varieties (French (FR), Belgian (BE), Swiss (CH), Quebecois (QC)), 2 sexes and 4 decades (50-59, 60-69, 70-79, 80-89 y.o.). In a forced-choice perception task, 13 young (22-31) and 13 old (70-95) Parisian French listeners estimated the speaker's age as one of the 4 age classes, listening to the production of one meaningful read sentence per speaker (10 syllables, normalized intensity, 16 speakers in test-retest).

'Chronological age' is a fairly reliable predictor of a speaker's 'perceived age' (Pearson correlation r=.75). However, we notice a tendency to overestimate the age of the speakers in the [50-59] group and underestimate the age of speakers from 70 y.o. (as found by others e.g. Huntley et al., 1987, Hunter et al., 2016). In other terms, speakers in the 60-69 y.o. group appear to sound closer to their chronological age, but this result is to consider with caution since there is a tendency for responses to converge to the central categories in these tasks. Differences between 'perceived' and 'chronological' age can be explained by different types of factors, linked either to regional and generational mismatch between listeners and speakers, or to the acoustic characteristics of the speakers' speech:

i) age estimations are found to depend on the shared origin between speaker and listener (cf. Foulkes et al. 2010, Nagao 2006, but Braun et Cerrato 1999): at a same chronological age, French speakers are rated by the Parisian French listeners as younger than speakers of other dialects (fig. 1). Links between 'perceived' and 'chronological' age also vary according to the dialectal groups: for Swiss speakers the correlation is higher (r=.87) than for other groups (.72 BE, .78 QC, .74FR).

ii) estimations are also found to depend on the age difference between the listeners and the speakers, as found by others (e.g. Lindville & Korabic 1986; Goy et al., 2016). Unexpectedly, the listeners' age matters more for the estimation of younger speakers: the age of speakers over 70 is equally underestimated by young and older listeners, while speakers under 70 are mainly judged older than their chronological age by the older listeners. This result does not reflect difficulties of the older listeners, whose intra-judge agreement is the same as young listeners.

iii) age estimation depends on acoustical characteristic of the speaker. Among the few temporal and spectral acoustic cues tested so far, speech rate appears to be the best predictor of the speakers' age, although it accounts for only a small part of age variation. Interestingly, speech rate better predicts perceived (r=-0.40) than chronological age (r=-0.25), see fig. 2. In line with Harnsberger et al. 2006 and production studies such as Ramig 1983, this result suggests that, at the same chronological age, speakers do not sound equally old based on their speech rate.

Overall, this study shows that a speaker's age can be estimated from his/her speech, and that this estimation depends both on the acoustical cues in the signal and on individual factors linked to the perceptual process of estimating personal information. Multiple cues indexing age are present in the speech signal and follow-up studies will go beyond the few acoustical cues analyzed here. Results also underline the potential lack of homogeneity within groups when classifying speakers by chronological age, rather than perceived age. This result has strong implications for the constitution of age-matched groups for the study of pathological speech for instance, but also for the understanding of sociophonetic variability.

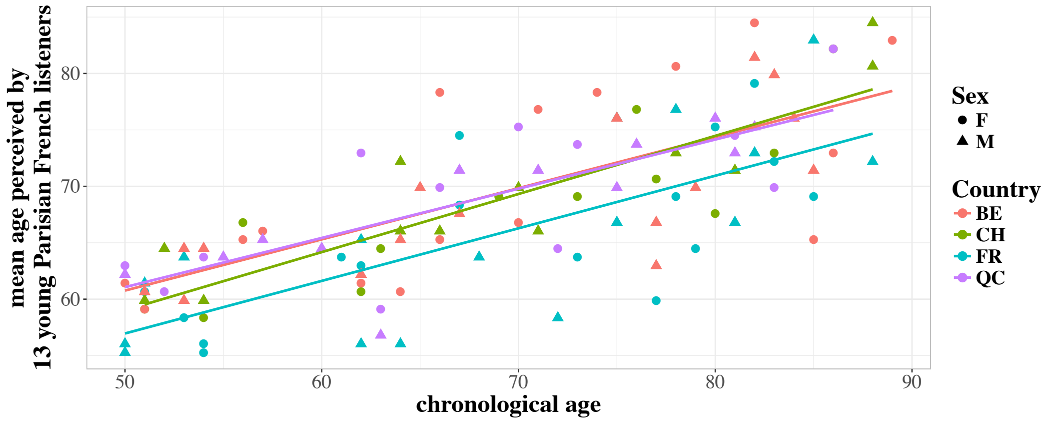


Figure 1. *Scatterplot of chronological age vs. age perceived by 13 young Parisian French listeners, for the 112 speakers distributed in 4 countries and balanced between F and M.*

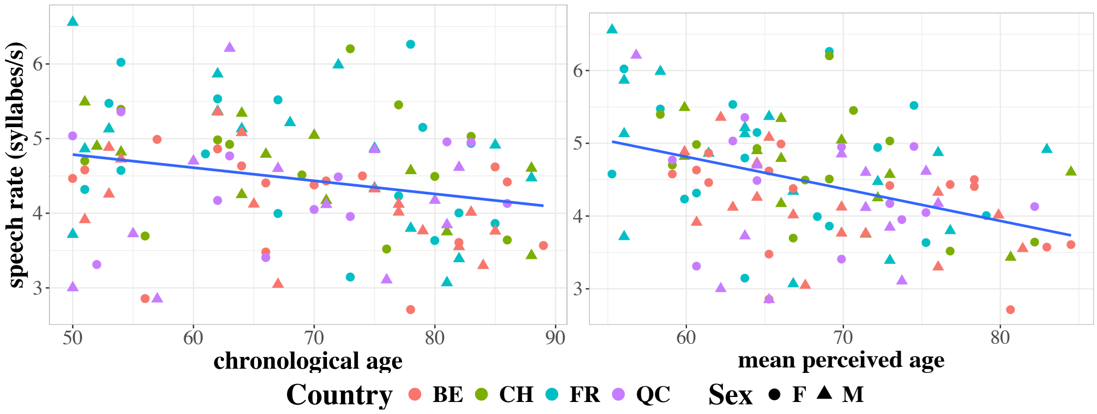


Figure 2. *Scatterplot of speech rate as a function of chronological age (left) or age perceived by 13 young Parisian French listeners (right) for the 112 speakers.*

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

[2] Foulkes, P., Docherty, G., Khattab, G., & Yaeger-Dror, M. (2010). Sound judgements: perception of indexical features in children's speech. In D. R. P. and N. Niedzielski (Ed.), *A Reader in Sociophonetics* (pp. 327–356). United States: De Gruyter Mouton.

[3] Eckert, P. (1997). Age as a sociolinguistic variable. In F. Coulmas (Ed.), *Handbook of sociolinguistics* (pp. 151–167). Oxford: Blackwell.

[4] Huntley, R., Holline, H., Shipp, T. (1987). Influences of listener characteristics on perceived age estimations. *Journal of Voice*, 1(1), 49-52.

[5] Hunter, E. J., Ferguson, S. H., & Newman, C. A. (2016). Listener estimations of talker age: A meta-analysis of the literature. *Logopedics Phoniatrics Vocology*, 41(3), 101-105.

[6] Nagao, K. (2006). *Cross-language study of age perception*. Doctoral dissertation, Department of Linguistics, Indiana University, Bloomington.

[7] Braun, A., & Cerrato, L. (1999). Estimating speaker age across languages. *Proceedings of ICPhS 1999*, 1369-1372.

[8] Linville, S. E., & Korabic, E. W. (1986). Elderly listeners' estimates of vocal age in adult females. *The Journal of the Acoustical Society of America*, 80(2), 692-694.

[9] Goy, H., Pichora-Fuller, M. K., & van Lieshout, P. (2016). Effects of age on speech and voice quality ratings. *The Journal of the Acoustical Society of America*, 139(4), 1648-1659.

[10] Harnsberger, J. D., Shrivastav, R., Brown, W. S. Jr., Rothman, H., & Hollien, H. (2008). Speaking rate and fundamental frequency as speech cues to perceived age. *J. voice*, 22(1), 58-69.

[11] Ramig, L. A. (1983). Effects of physiological aging on speaking and reading rates. *Journal of communication disorders*, *16*(3), 217-226.

Prominence and Boundary are two distinct phenomena in French: perceptual evidence

[1]Roxane Bertrand, [2]Corine Astésano & [1]Noël Nguyen
[1]Laboratoire Parole et Langage (CNRS UMR 7309) & Aix-Marseille Université ; [2]Octogone-Lordat, Université de Toulouse, UT2J, Toulouse, France

French prosodic characteristics are particularly challenging for phonological theories. Despite a wide body of literature, some central issues are still discussed such as the existence of stress at the word level, the existence of one or two pitch accents and at which level of prosodic constituency they surface [1]. Current models of French accentuation unanimously consider French accentuation as being post-lexical, with a primary final accent (FA) and a secondary initial accent (IA) marking the *ap* level. There is, however, no clear consensus as to the respective status of both FA and IA. Whereas [2] considers IA as a pitch accent like FA (H*), most authors describe IA (LHi) as a 'loose boundary marker' because its peak can be aligned with up to the third syllable of a word in a long *ap* [3; 4]. Only the L tone of LHi is consistently aligned with the beginning of the lexical word. IA is also secondary insofar as it is said to yield to FA in case of tonal crowding on shorter constituents, and essentially has a rhythmic function [3; 5]. Recent accounts however show speakers' use of IA as a more consistent marker of prosodic structure than FA, and quite independent from rhythmic constraints [6]. The status of FA is not entirely clarified either: although it is clearly a *pitch accent* at the *ap* level for most current models, some descriptions suggest that it may survive at higher levels of prosodic constituency [2: LH*-H%; 5], while the most widespread view suggests that it looses its metric quality at the *IP* level in favour of the sole boundary tone H% [7; 3]. This latter proposition stems from a phonological phenomenon quite specific to French, *i.e.* the syncretism between accent (LH*) and intonation contours (L% or H%), which blurs the clear acoustic realization of FA. Because stress is also not lexically distinctive, it has lead to the qualification of French as a 'boundary language' [8; 9] at the post-lexical level, or even as a 'language without accent' [7].

In the present study, perception is used to help shed light on those core issues by accounting for those prosodic parameters that are actually processed by listeners. It also helps answer some issues that are difficult to account for by the sole tonal annotation of the speech signal. As exemplified by [10], calling upon the postlexicality of French accentuation, it is well believed that French listeners are "deaf" to prominence. The present study aims at more specifically testing the ability of French listeners to both perceive *and* distinguish prominence and boundaries. We hypothesize that listeners are capable of perceiving different levels of boundaries and prominence, and are able to dissociate these two phonological phenomena, even in the case of syncretism between accentuation and intonation. We also test the level of prosodic structure preferentially marked by IA and/or FA.

The perception experiment was carried out on a corpus of syntactically ambiguous sentences that can be disambiguated via prosodic cues (sub-corpus taken from [6]). Syntactic ambiguity is created by manipulating the adjective scope (low or high syntactic attachment of the adjective *A* to one or two Nouns *N1* and *N2*), yielding 4 prosodic sites and 3 prosodic boundary strengths: *w*-boundary; *ap*-boundary and *ip*-boundary (see **Figure 1**). The prosodic structure is also manipulated with regards to constituents' length (one to four syllables), resulting in 16 original scripts. 32 sentences uttered by one female speaker were used for the present experiment. 18 naive listeners had to perform two separate perception tasks (counterbalanced between listeners): a task where they had to evaluate the level of *boundary* between the words, on a scale from 0 to 3; a task where they had to evaluate the level of *prominence* on each syllable of the sentences, on a scale from 0 to 3. Ordinal logistic mixed models [11] were used to account for 1) the perception of IA and FA prominences; 2) the perception of boundary strengths; 3) the links between boundary and prominence perception.

Results indicate that 1) both IA and FA prominence are perceived as significantly more salient than surrounding syllables, with a stronger perception of IA than FA, and as early as the word level. Moreover, FA at *ip* boundaries is perceived as metrically strong despite the syncretism between accentuation and intonation contours, indicative of a phonological representation of stress. These two results question the notion of stress deafness, at least in the native linguistic system (**Figure 2**). 2) Prosodic boundaries' strengths are not perceived as predicted by syntactic structure. Rather, if stronger (*ip*) boundaries are indeed perceived as stronger than *w* and *ap* boundaries, the N1-*ip*-N2 is perceived as stronger than N2-*ip*-A. Also, *w* and *ap* boundaries are perceived as equally strong. These results will be further discussed with regards to the literature on prosodic constituency in French. 3) Dissociation between boundary and prominence is observed insofar as the same prominence score can be found in association with any of the 3 boundary strengths. More precisely, IA is always perceived with the same strength independently from boundary strength, while FA is also perceived with the same strength at the *w*, the *ap* and the N2-*ip*-A level. It only is correlated with boundary strength at the N1-*ip*-N2 boundary site. Altogether, these results indicate that FA *and* IA are equally important markers of the prosodic structure, and that both accents surface in turn or concomitantly to mark prosodic constituency, potentially as early as the prosodic word level [12], just below the *ap* level. They thus question both the notions of stress deafness and post-lexicality in French. They will be discussed in depth also with regards to ongoing acoustic and neuroimaging data.

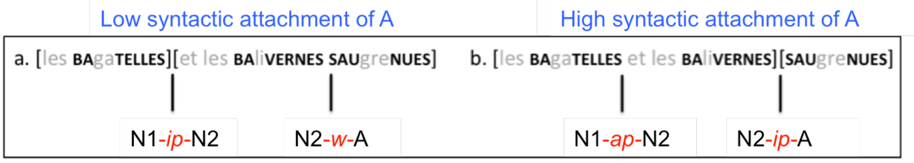


Figure 1: *The 4 prosodic sites of interest on trisyllabic words in 'Les bagatelles et les balivernes saugrenues' ('Crazy trifles and nonsense') in both syntactic conditions. Bold syllables indicate where FA and IA can potentially occur to mark prosodic structure.*

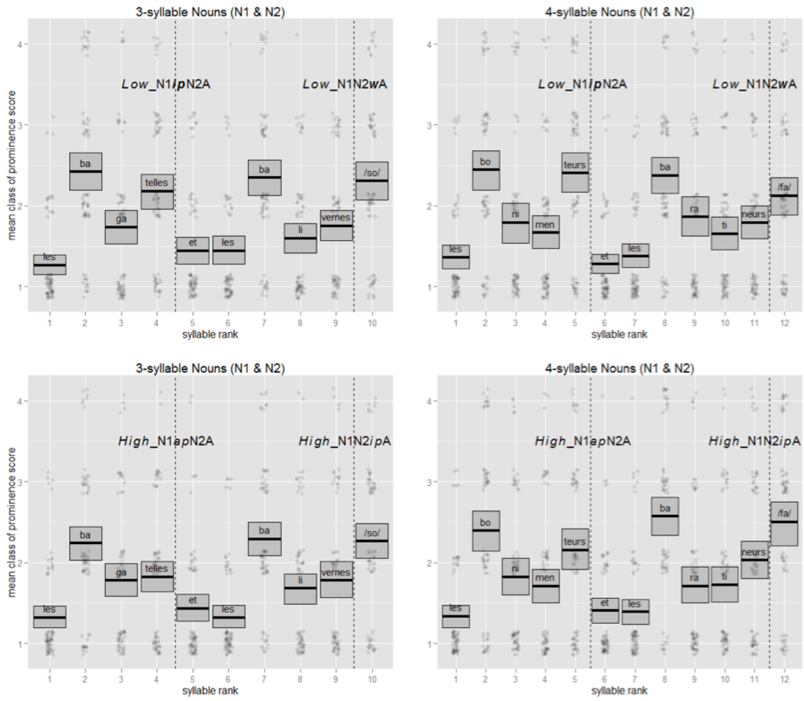


Figure 2: *Perception score of syllables in both high (bottom) and low (top) syntactic attachment conditions for three- (left) and four-syllable (right) words N1s and N2s (with A length collided). Boxes indicate the confidence interval of the mean score. Dots indicate individual responses (with jitter). Dotted lines indicate the boundary sites. Left: Syllables 2 and 7 show perception of IA; syllables 4 and 9 show perception of FA. Right: Syllables 2 and 8 show perception of IA; syllables 5 and 11 show perception of FA.*

References

[1] 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.

[2] Post, B. 2000. *Tonal and Phrasal Structures in French Intonation*. Thesus, The Hague.

[3] Jun, S.-A. ; Fougeron, C. 2000. A phonological model of French intonation. In: Botinis, A. (Ed.). *Intonation: Analysis, Modelling and Technology*. Kluwer, Boston: 209–242.

[4] Welby, P. 2003. *The slaying of Lady Mondegreen, being a study of French tonal association and alignment and their role in speech segmentation*. PhD dissertation, OSU.

[5] Di Cristo, A. 2000. Vers une modélisation de l'accentuation en français. *Journ. of French Lang. Studies, 10,* 27-44.

[6] Astésano, C.; Bard, E.; Turk, A. 2007. Structural influences on Initial Accent placement in French. Lang. and Speech, 50 (3), 423-446.

[7] Rossi, M. 1980. Le français, langue sans accent ? In I. Fónagy & P. Léon: *L'accent en français contemporain*, *15*, 13-51.

[8] Vaissière, J. 1990. Rhythm, accentuation and final lengthening in French. In : J. Sundberg, L. Nord & R. Carlson. *Music, language, speech and brain*: 108-120.

[9] Beckman, M.E. 1992. Evidence for Speech Rhythms across Languages. In Tohkura; Vatikiotis-Bateson; Sagisaka (eds.) : Speech Perception, Production and Linguistic Structure. Tokyo: 457-463.

[10] Dupoux, E., Pallier, C., Sebastian, N., & Mehler, J. (1997). A destressing "deafness" in French? *Journal of Memory and Language, 36(3),* 406-421.

[11] Christensen, R. H. B. (2012). Regression Models for Ordinal Data R package version 2012.09-11

[12] Selkirk, E. (1996). The prosodic structure of function words. In: K. Demuth and J. Morgan: *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*: 187-213.

Prominence and Boundary are two distinct phenomena in French: perceptual evidence

[1]Roxane Bertrand, [2]Corine Astésano & [1]Noël Nguyen
[1]Laboratoire Parole et Langage (CNRS UMR 7309) & Aix-Marseille Université ; [2]Octogone-Lordat, Université de Toulouse, UT2J, Toulouse, France

French prosodic characteristics are particularly challenging for phonological theories. Despite a wide body of literature, some central issues are still discussed such as the existence of stress at the word level, the existence of one or two pitch accents and at which level of prosodic constituency they surface [1]. Current models of French accentuation unanimously consider French accentuation as being post-lexical, with a primary final accent (FA) and a secondary initial accent (IA) marking the *ap* level. There is, however, no clear consensus as to the respective status of both FA and IA. Whereas [2] considers IA as a pitch accent like FA (H*), most authors describe IA (LHi) as a 'loose boundary marker' because its peak can be aligned with up to the third syllable of a word in a long *ap* [3; 4]. Only the L tone of LHi is consistently aligned with the beginning of the lexical word. IA is also secondary insofar as it is said to yield to FA in case of tonal crowding on shorter constituents, and essentially has a rhythmic function [3; 5]. Recent accounts however show speakers' use of IA as a more consistent marker of prosodic structure than FA, and quite independent from rhythmic constraints [6]. The status of FA is not entirely clarified either: although it is clearly a *pitch accent* at the *ap* level for most current models, some descriptions suggest that it may survive at higher levels of prosodic constituency [2: LH*-H%; 5], while the most widespread view suggests that it looses its metric quality at the *IP* level in favour of the sole boundary tone H% [7; 3]. This latter proposition stems from a phonological phenomenon quite specific to French, *i.e.* the syncretism between accent (LH*) and intonation contours (L% or H%), which blurs the clear acoustic realization of FA. Because stress is also not lexically distinctive, it has lead to the qualification of French as a 'boundary language' [8; 9] at the post-lexical level, or even as a 'language without accent' [7].

In the present study, perception is used to help shed light on those core issues by accounting for those prosodic parameters that are actually processed by listeners. It also helps answer some issues that are difficult to account for by the sole tonal annotation of the speech signal. As exemplified by [10], calling upon the postlexicality of French accentuation, it is well believed that French listeners are "deaf" to prominence. The present study aims at more specifically testing the ability of French listeners to both perceive *and* distinguish prominence and boundaries. We hypothesize that listeners are capable of perceiving different levels of boundaries and prominence, and are able to dissociate these two phonological phenomena, even in the case of syncretism between accentuation and intonation. We also test the level of prosodic structure preferentially marked by IA and/or FA.

The perception experiment was carried out on a corpus of syntactically ambiguous sentences that can be disambiguated via prosodic cues (sub-corpus taken from [6]). Syntactic ambiguity is created by manipulating the adjective scope (low or high syntactic attachment of the adjective *A* to one or two Nouns *N1* and *N2*), yielding 4 prosodic sites and 3 prosodic boundary strengths: *w*-boundary; *ap*-boundary and *ip*-boundary (see **Figure 1**). The prosodic structure is also manipulated with regards to constituents' length (one to four syllables), resulting in 16 original scripts. 32 sentences uttered by one female speaker were used for the present experiment. 18 naive listeners had to perform two separate perception tasks (counterbalanced between listeners): a task where they had to evaluate the level of *boundary* between the words, on a scale from 0 to 3; a task where they had to evaluate the level of *prominence* on each syllable of the sentences, on a scale from 0 to 3. Ordinal logistic mixed models [11] were used to account for 1) the perception of IA and FA prominences; 2) the perception of boundary strengths; 3) the links between boundary and prominence perception.

Results indicate that 1) both IA and FA prominence are perceived as significantly more salient than surrounding syllables, with a stronger perception of IA than FA, and as early as the word level. Moreover, FA at *ip* boundaries is perceived as metrically strong despite the syncretism between accentuation and intonation contours, indicative of a phonological representation of stress. These two results question the notion of stress deafness, at least in the native linguistic system (**Figure 2**). 2) Prosodic boundaries' strengths are not perceived as predicted by syntactic structure. Rather, if stronger (*ip*) boundaries are indeed perceived as stronger than *w* and *ap* boundaries, the N1-*ip*-N2 is perceived as stronger than N2-*ip*-A. Also, *w* and *ap* boundaries are perceived as equally strong. These results will be further discussed with regards to the literature on prosodic constituency in French. 3) Dissociation between boundary and prominence is observed insofar as the same prominence score can be found in association with any of the 3 boundary strengths. More precisely, IA is always perceived with the same strength independently from boundary strength, while FA is also perceived with the same strength at the *w*, the *ap* and the N2-*ip*-A level. It only is correlated with boundary strength at the N1-*ip*-N2 boundary site. Altogether, these results indicate that FA *and* IA are equally important markers of the prosodic structure, and that both accents surface in turn or concomitantly to mark prosodic constituency, potentially as early as the prosodic word level [12], just below the *ap* level. They thus question both the notions of stress deafness and post-lexicality in French. They will be discussed in depth also with regards to ongoing acoustic and neuroimaging data.

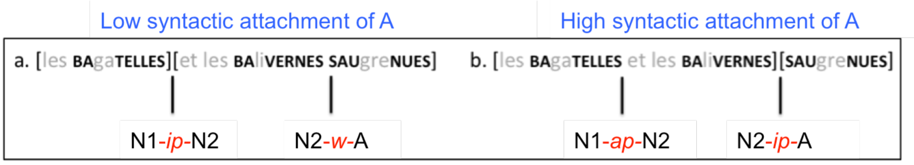


***Figure 1**: The 4 prosodic sites of interest on trisyllabic words in 'Les bagatelles et les balivernes saugrenues' ('Crazy trifles and nonsense') in both syntactic conditions. Bold syllables indicate where FA and IA can potentially occur to mark prosodic structure.*

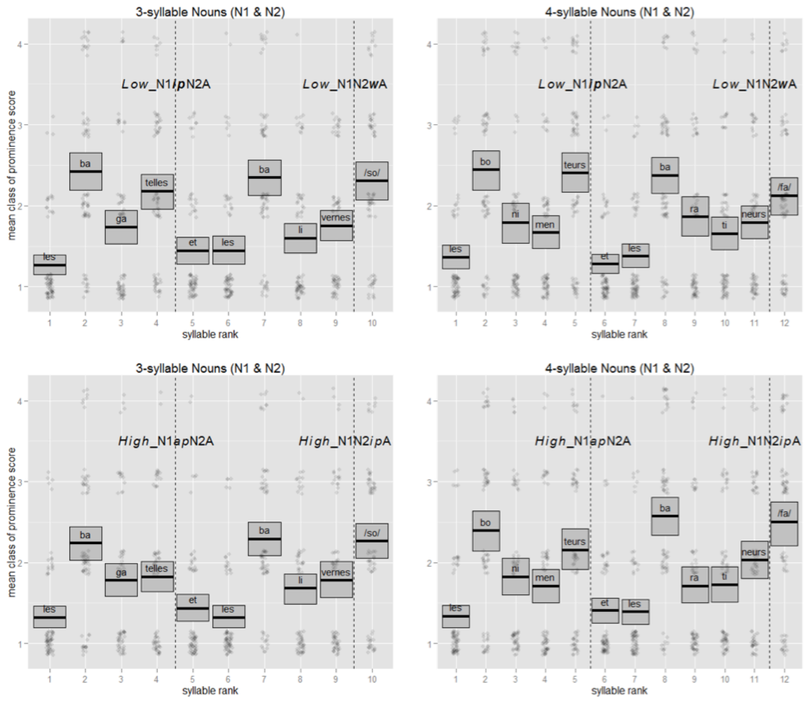


***Figure 2**: Perception score of syllables in both high (bottom) and low (top) syntactic attachment conditions for three- (left) and four-syllable (right) words N1s and N2s (with A length collided). Boxes indicate the confidence interval of the mean score. Dots indicate individual responses (with jitter). Dotted lines indicate the boundary sites. Left: Syllables 2 and 7 show perception of IA; syllables 4 and 9 show perception of FA. Right: Syllables 2 and 8 show perception of IA; syllables 5 and 11 show perception of FA.*

References

[1] 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.
[2] Post, B. 2000. *Tonal and Phrasal Structures in French Intonation*. Thesus, The Hague.
[3] Jun, S.-A. ; Fougeron, C. 2000. A phonological model of French intonation. In: Botinis, A. (Ed.). *Intonation: Analysis, Modelling and Technology*. Kluwer, Boston: 209–242.
[4] Welby, P. 2003. *The slaying of Lady Mondegreen, being a study of French tonal association and alignment and their role in speech segmentation*. PhD dissertation, OSU.
[5] Di Cristo, A. 2000. Vers une modélisation de l'accentuation en français. *Journ. of French Lang. Studies, 10,* 27-44.
[6] Astésano, C.; Bard, E.; Turk, A. 2007. Structural influences on Initial Accent placement in French. Lang. and Speech, 50 (3), 423-446.
[7] Rossi, M. 1980. Le français, langue sans accent ? In I. Fónagy & P. Léon: *L'accent en français contemporain*, *15*, 13-51.
[8] Vaissière, J. 1990. Rhythm, accentuation and final lengthening in French. In : J. Sundberg, L. Nord & R. Carlson. *Music, language, speech and brain*: 108-120.
[9] Beckman, M.E. 1992. Evidence for Speech Rhythms across Languages. In Tohkura; Vatikiotis-Bateson; Sagisaka (eds.) : Speech Perception, Production and Linguistic Structure. Tokyo: 457-463.
[10] Dupoux, E., Pallier, C., Sebastian, N., & Mehler, J. (1997). A destressing "deafness" in French? *Journal of Memory and Language, 36(3),* 406-421.
[11] Christensen, R. H. B. (2012). Regression Models for Ordinal Data R package version 2012.09-11
[12] Selkirk, E. (1996). The prosodic structure of function words. In: K. Demuth and J. Morgan: *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*: 187-213.

Icelandic question intonation

Nicole Dehé and Bettina Braun
University of Konstanz

This presentation reports on the first experimental study investigating question intonation in Modern Icelandic. Specifically, we carried out a production experiment to investigate the intonation of polar questions and constituent questions (henceforth *wh*-questions). In Icelandic, polar questions are verb-initial (V1) (see (1)). *Wh*-questions (see (2)) are formed with *wh*-interrogative pronouns (Icelandic *hv*-pronouns). The canonical position of the *hv*-pronoun or *hv*-phrase is sentence-initial, i.e. in the position preceding the finite verb.

According to previous literature based mostly on introspective data, the default boundary tone for all sentence types is L%. Icelandic neutral declaratives are typically realized with a nuclear L+H* pitch accent, followed by a fall to L%. If the nuclear tune is preceded by prenuclear prominence, the typical accent is a late rise (L*+H) (see [1], [2]). According to [2], the typical nuclear pitch accent in Icelandic polar questions is a rise from a low target on the accented syllable (late rise; L*+H), the peak being reached after the end of the nuclear syllable. Combined with the low boundary tone, the typical nuclear contour is thus a rising-falling one (L*+H L%). The intonation of Icelandic polar questions thus differs crucially from polar questions in related languages, which often end in a rise to H%, e.g. in English ([3-6]). According to [7], the typical contour associated with a *wh*-question starts high, followed by a H* nuclear accent and an L% boundary tone. Nuclear pitch accent types, but not boundary tones, thus distinguish between sentence types in Icelandic.

We carried out a controlled production experiment testing the following hypotheses: (a) Both polar and wh-questions are typically produced with an intonational contour falling to L%. (b) The two question types differ in nuclear pitch accent type. Specifically, polar questions are typically realised with a late rise (L*+H), while *wh*-questions are typically realised with a monotonal H* pitch accent. (c) The typical contour associated with *wh*-questions starts high.

155 polar questions and 162 *wh*-questions, produced by 17 speakers (aged 22-32; average 26.9; 6 male), entered the analysis. The data were annotated by the first author. A sample of 15% of the data was additionally annotated by the second author; disagreements were resolved. The results largely confirm the observations made in previous literature, but they also add important insights.

(a) The default boundary tone is indeed L% across question types (96.1% in polar questions, 87% in *wh*-questions).

(b) The two question types differ in nuclear pitch accent type. While polar questions are mostly realized with bitonal rising nuclear accents (late rise: 56.1%; L*+H: 44.5%, L*+^H: 11.6%), *wh*-questions typically have monotonal peak accents (62.3% overall, combining H*, !H*, ^H*). Monotonal pitch accents are very rare in polar questions (2%). Both question types are also frequently realised with an early rise nuclear accent (L+H*, L+^H*, L+!H*), which suggests that the late rise is a kind of neutral accent is Icelandic. It is typical in declaratives, but it also has the capacity to mark questionhood, even though in questions it is not the most frequent one.

(c) *Wh*-questions indeed often start high (71% in *wh*-questions vs. 12% in polar questions). The high start in *wh*-questions is due to either an initial %H boundary tone, which has not previously been identified as part of the Icelandic tonal inventory, or to a H* prenuclear accent associated with the *wh*-pronoun.

To summarize, the most important results are that (a) the boundary tone does not distinguish between sentence types anywhere, (b) sentence types are distinguished by nuclear accent types, although the early rise (L+H*) is possible across the board, and (c) the prenuclear area contributes to the intonational distinction between question types.

(1)	Át	álfur-inn	ost-inn?
	ate	elf-DEF.NOM	cheese.ACC-DEF.ACC
	'Did the elf eat the cheese?'		

(2)	*Hver*	hefur	gefið	Maríu	þennan	hring?
	Who	has	given	Mary	this	ring

[1]N. Dehé, "The nature and use of Icelandic prenuclear and nuclear pitch accents: Evidence from F0 alignment and syllable/segment duration.," *Nordic Journal of Linguistics,* vol. 33, pp. 31-65, 2010.

[2]K. Árnason, *The Phonology of Icelandic and Faroese*. Oxford: Oxford University Press, 2011.

[3]C. Bartels, *The Intonation of English Statements and Questions. A Compositional Interpretation.* New York & London: Garland Publishing, 1999.

[4]N. Hedberg, J. M. Sosa, and E. Görgülü, "Early and Late Nuclei in Yes-No Questions: Tails or High Rises?," *Proceedings of Speech Prosody 2008, Campinas, Brazil, May 6-9, 2008,* pp. 229-232, 2008.

[5]N. Hedberg and J. M. Sosa, "The prosody of questions in natural discourse," in *Proceedings of the First International Conference on Speech Prosody (Speech Prosody 2002), Aix-en-Provence, France.* , 2002, pp. 375-378.

[6]J. Pierrehumbert and J. Hirschberg, "The meaning of intonational contours in the interpretation of discourse," in *Intentions in Communication*, P. R. Cohen, J. Morgan, and M. E. Pollack, Eds., ed Cambridge, MA: MIT Press, 1990, pp. 271-311.

[7]K. Árnason, *Hljóð. Handbók um hljóðfræði og hljóðkerfisfræði.* vol. 1. Bindi. Reykjavík: Almenna bókafélagið, 2005.

The realization of focus in monolingual and bilingual Spanish

Maria del Mar Vanrell[1] & Ingo Feldhausen[2]
[1]Universitat de les Illes Balears, [2]Goethe-Universität Frankfurt

In this paper, we shed new light on the question of how narrow contrastive and information focus is realized in Central-Peninsular Spanish. We show that there is an important difference with respect to the use of both pitch accents and syntactic strategies for realizing focus between monolingual native speakers and German-Spanish bilingual speakers. Our five main findings:

(a) Bilingual speakers realize both types of focus almost always by stress shift, (1a), and the pitch accent is predominantly realized by L+H* (see section *Bilinguals*);
(b) Monolingual speakers, in turn, realize information focus by different strategies (*cf.* (1b), (2), and (3)), but stress shift is not a relevant option (see section *No stress shift*);
(c) Cleft constructions are used by monolinguals for both focus types even though there are certain preferences (see section *Cleft and focus type*);
(d) Focus does not have to bear always sentential stress: in clefts, prosodic alignment can be a sufficient correlate of focus (see section *Focus without sentential stress*);
(e) Monolinguals typically realize the pitch accents by L+H* for non-final focused constituents and L* for final focused constituents.

We further argue that existing discrepancies between claims made in theoretical work on the one hand and in empirical work on the other can often be reduced to diatopic differences.

Methodology: We conducted a production test based on semi-spontaneous speech designed to elicit different focus readings (narrow informational and contrastive focus on the subject and (in)direct objects) by means of question-answer pairs from short picture stories. A total of 2508 contours were obtained (Monolinguals: 1848 = 24 short stories x 11 questions x 7 speakers; bilinguals: 660 = 12 x 11 x 5; all native speakers of Central-Peninsular Spanish).

No stress shift: There is an ongoing discussion on how focus is realized in Spanish. Theoretical work (such as Zubizarreta 1998, Gutiérrez-Bravo 2002) argues that neutrally focused elements must be located in sentence-final position (via *p-movement*, (1b)) in order to receive main stress by means of the *Nuclear Stress Rule*. Empirical studies, in turn, show that neutrally focused elements actually can be realized *in situ* (1a) and that this option reflects the predominant strategy for focus realization (e.g. Muntendam 2013, Leal et al. 2018, among many others). Our empirical results of the monolingual speakers (N=7) show that stress shift is not an option in Central-Peninsular Spanish and suggest that dialectal variation must be taken into account as a decisive factor involved in the variation of focus realization strategies.

Cleft and focus type: While the cleft constituent (such as *Juan* in (3)) is generally considered to be the contrastively focused element in Spanish (see, e.g., Zubizarreta 1998), Moreno Cabrera (1999: 4298f.) states that simple clefts on the one hand and (inverted) pseudo-clefts on the other hand have different information structural properties. Our study – as far as we know – represents the first empirical verification of this claim and confirms it; see Table 2.

Focus without sentential stress: It is generally accepted that focus in Spanish bears sentential stress (see, e.g., Ortiz-Lira 1994, Zubizarreta 1998 among many others). However, contrary to what has been claimed in the past, our results show the contrastively focused constituent in clefts such as (3a) does not always bear sentential stress (in up to 80% of the cases) – independently of the grammatical function of the clefted element.

Bilinguals: We tested five early bilinguals (so far) who grew up and still live in Germany and who speak Central-Peninsular Spanish as a heritage language. The speakers show a clear preference for stress shift in both informational and contrastive focus, see Table 1 (in line with other studies on bilinguals, e.g. Leal et al. 2018). The realized focal pitch accent is almost always L+H*, but it is longer and more intense in contrastive contexts. Interestingly, the few instances of p-movement attested in the bilingual data occur with contrastive focus and not with information focus. Thus, the bilinguals clearly differ from the monolinguals. Future research will show whether the differences might be due to the influence of German (a language allowing for stress shift) or whether stress shift is a default strategy of bilinguals.

(1) a. [$_F$ Los aLUMnos] se enfrentaron con la policía. (*Europ. Sp. / LatinAmSp.)
‘The students confronted the police’.
b. Se enfrentaron con la policía [$_F$ los aLUMnos]. (Europ. Sp. / ^{ok}LatinAmSp.)

(2) [$_{CF}$ ManZAnas] compró Pedro (y no peras). Contrastive focus fronting
‘Pedro eats apples (and not pears).’

(3) a. Es *Juan* el que viene. Clefts
‘It is Juan who comes.’
b. El que viene es *Juan*. Pseudo-clefts
c. *Juan* es el que viene. Inverted pseudo-clefts

Information focus		
	Monolinguals	Bilinguals
[$_F$S]	Clefting 71.1% P-movement 14.5%	Stress shift 77% Clefting 18%
[$_F$O$_{OD}$]	P-movement 47.9% Clefting 23.3%	Stress shift 83% Clefting 15%
[$_F$O$_{OI}$]	Neutral WO 43.6 % Clefting 21.3%	Neutral WO 99%

Contrastive focus		
	Monolinguals	Bilinguals
[$_{CF}$S]	Clefting 61.4% Focus fronting 15%	Stress shift 72% Clefting 23%
[$_{CF}$O$_{OD}$]	Clefting 61.8% Focus fronting 23.6%	Stress shift: 63% P-movement: 27%
[$_{CF}$O$_{OI}$]	Clefting 41.2% Focus fronting 23.7%	Neutral WO 87%

Table 1: Types and frequency of focus marking strategies in neutral focus (left panel) and contrastive focus (right panel) declaratives; types of clefts (see (3)) are not distinguished here.

Neutral focus	Clefts 44,9% Pseudo-clefts 13,4 % Inverted pseudo-clefts 41,5%
Contrastive focus	Clefts 70,98% Pseudo-clefts 23,52 % Inverted pseudo-clefts 5,4 %

Table 2: Types and frequency of cleft constructions attested in neutral and contrastive focus declaratives (monolingual speakers).

References:

[1] Gutiérrez-Bravo, R. 2002. Focus, Word Order variation and intonation in Spanish and English. In C. Wiltshire & J. Camps (eds), *Romance Phonology and Variation*. John Benjamins, 39-53.

[2] Leal, T., Destruel, E. & Hoot, B. 2018. The realization of information focus in monolingual and bilingual native Spanish. *LAB* 8(2): 217–251

[3] Moreno Cabrera, J.C. 1999. Las funciones informativas: las perífrasis de relativo. In: *Gramática Descriptiva de la Lengua Española*, I. Bosque & V. Demonte (eds.), Madrid, Espasa, 4245-4302.

[4] Muntendam, A. 2013. On the nature of cross-linguistic transfer. *BLC* 16(1), 111–131.

[5] Ortiz-Lira, H. 1994. A contrastive analysis of English and Spanish sentence accentuation. PhD dissertation, U of Manchester.

[6] Zubizarreta, M.L. 1998. *Prosody, Focus, and Word Order*. MIT Press.

Information seeking vs. rhetorical questions: from gradience to categoricity

Elisabeth Delais-Roussarie[1] and Claire Beyssade[2]
[1]*Université de Nantes - UMR 6310 (LLING),* [2]*Université Paris 8 – UMR 7023 (SFL)*

It is well established that declarative sentences, which generally convey an assertion as in (1a), can be used for questioning as in (1b), often labelled as a declarative question. But conversely, interrogative sentences, usually interpreted as information seeking question (ISQ), can convey an assertive content, as illustrated by their use as rhetorical questions (RhQ). See the sentence (3), whose interpretation varies depending on the context in which it occurs ((2a) *vs.* (2b)). There is already quite a large number of studies that provide an analysis of the prosodic cues that allow turning declarative sentences into questions (a.o. [1], [2], [3], [4]), but there are only few studies on the prosodic cues distinguishing ISQs from a RhQs (see [5], [6], [7] on German, and [8] and [9] on English and Icelandic, all using the same protocol).

Our aim here is to investigate the difference between ISQs and RhQs in French. This is interesting for at least two reasons: (i) there is almost no study on the prosody of rhetorical questions in French ([10]); and (ii) French is a romance language in which the form of questions varies a lot and relatively freely (fronted *vs.*in-situ wh-question, declarative vs. interrogative polar questions).[1]

For the study, the protocol used by Braun and colleagues ([5] to [9]) was reduplicated. We constructed 21 pairs of *wh*- and polar questions, each of them occurring in two contexts: one triggering an ISQ interpretation (2a) for the uttered question (3), and another one constraining the question to be interpreted as a RhQ (2b). Twelve monolingual native speakers of French participated. After a visual display of the context, they had to read aloud the target sentence in a way suitable to the given context.

An acoustic analysis was done in order to evaluate which prosodic features contribute to the interpretation of such interrogative sentences. Four distinct features are taken into account: form of the final contour, speech rate, pitch range (over the whole sentence and over the specifically questioned part, i.e. wh-word in wh-question), realisation of the wh-word or the specifically questioned part in polar question.

Concerning polar questions, preliminary results show that (i) speech rate is always faster in ISQs than in RhQs; (ii) a rising contour H*H% is observed at the end of ISQs in 83% of the cases, whereas three distinct final contours are observed in RhQs; and (iii) pitch range is larger in ISQs than in RhQs in more than 90% of the case. In addition, concerning the specifically questioned part of questions (like *quelqu'un 'anyone'* in (4)), its duration is shorter and its register greater in ISQs than in the corresponding RhQs, despite its non-final position. As for wh-questions, the analysis of the data shows that (i) speech rate is always faster except for one speaker; (ii) ISQs end with a rising contour in 80% of the cases, whereas a falling contour (H*L%) or a plateau are observed in RhQs (see fig. 1); (iii) F0 key is higher in ISQs than in RhQs, and pitch range is greater in more than 80% of the cases. Moreover, the wh-word is realized at a higher pitch in ISQs than in RhQs, and its duration is shorter in ISQs.

These results clearly show that ISQ and RhQ are realized differently. Note however that the differences are already occurring in the prenuclear domain (see also [11]). In addition, they are often related to gradient phenomena that have been considered as conveying meaning in alternative questions (see [12], difference in pitch register). Even if these results need to be validated against the entire corpus and by a perception experiment, they raise the question of the relationship between gradience and categoricity. Can a sum of gradient phenomena be the source of a categorical judgment? Can a gradient phenomenon convey linguistic meaning?

---

[1] In French, almost all declarative sentences, with the appropriate prosody, can convey questions. But these questions can only be ordinary (not rhetorical).

(1a) Il pleut. L% ‘It is raining’ (1b) Il pleut? H% ‘It is raining’

(2a) Context A triggering an ISQ interpretation:
You want to cook a dish with spinach for dinner. But your son has invited friends, and you don’t know whether they like this vegetable and will eat it or not. You say to the guests:

(2b) Context B triggering an RhQ interpretation:
In the canteen, for lunch, oven baked spinach dish is proposed. However, you know that nobody likes this disgusting vegetable. You say to your friends:

(3) Qui mange des épinards? ‘Who eats spinach?’

(4) Est-ce que quelqu’un mange des épinards ? ‘Does anyone eat spinach?’

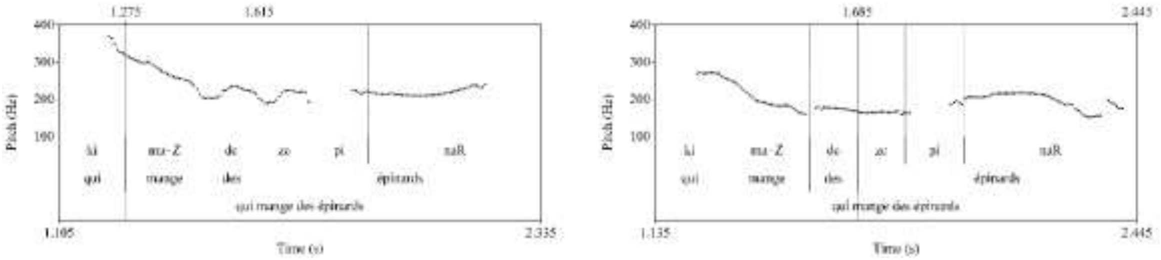


Figure 1: Pitch curve associated with the sentence *Qui mange des épinards* interpreted as a ISQ (left panel) and a RhQ (right panel).

[1] Bartels, C. 1999. *The intonation of English statements and questions: a compositional interpretation*. New York / London: Garland Publishing Inc

[2] Săfárŏvá, M. & Swerts, M. 2004. On recognition of declarative questions in English. *Proceedings of Speech Prosody 2004*, Nara, Japan.

[3] Vion, M. & Colas, A. 2006. Pitch cues for the recognition of yes – no questions in French. *Journal of Psycholinguistic Research* 35, 427–45.

[4] Frota, S., Butler, J. & Vigário, M. 2014. Infants’ perception of intonation: Is it a statement or a question? *Infancy* 19, 194–213.

[5] Neitsch, J., Braun, B. & Dehé, N. 2018. The role of prosody for the interpretation of rhetorical questions in German. *Paper presented at the 9th International Conference on Speech Prosody*, Poznan, Poland.

[6] Wochner, D., Schlegel, J., Braun, B. & Dehé, N. 2015. The prosody of rhetorical questions in German. *Proceedings of Interspeech 2015,* Dresden, Germany.

[7] Braun, B., Dehé, N., Neitsch, J., Wochner, D. & Zahner, K. To appear. The prosody of rhetorical and information-seeking questions in German. *Language and Speech.*

[8] Dehé, N., Braun, B. & Wochner, D. 2018. The prosody of rhetorical vs. information-seeking questions in Icelandic. *Proceedings of Speech Prosody*, Poznan, Poland. 403-407.

[9] Dehé, N. 2018. The prosody of rhetorical questions. In: *NELS 48: Proceedings of the Forty-Eighth Annual Meeting of the North East Linguistic Society*: Volume 1, ed. by Sherry Hucklebridge and Max Nelson, GLSA Publications.

[10] Di Cristo, A. 2016. *Les musiques du français contemporain: Essais sur l’accentuation, la métrique, le rythme, le phrasé prosodique et l’intonation du français contemporain*. Berlin/ Boston: De Gruyter

[11] Petrone, C. & Niebuhr, O. 2014. On the intonation of German intonation questions: The role of the prenuclear region. *Language and Speech* 57(1), 108–146

[12] Delais-Roussarie, E. & T. Turco (accepted). Intonation of Alternative Constructions in French: which cues allow distinguishing statements from questions? In *Romance Languages and Linguistic Theory 16: Selected papers from ‘Going Romance’ 30,* ed. by I. Feldhausen, I. Kuchenbrandt, M. Elsig & M. Neuhaus, John Benjamins Publishing Company.

Theory-Description-Theory: A round trip in French sign language phonology

Carlo Geraci (CNRS, IJN) & Justine Mertz (Unive. Paris-Diderot)

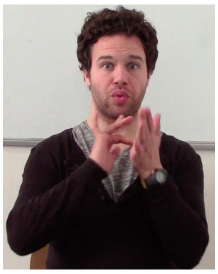
Goals. We use French sign language (LSF) data to evaluate descriptive and explanatory adequacy of current models of Sign Language (SL) phonology. Specifically, we show that LSF poses several problems for current theories of orientation both in terms of feature inventory and in terms of general accounts of orientation as a relative relation between the hand and the plane of articulation.

Background. SLs are natural languages that are perceived visually (vs. acoustically) and produced gesturally (vs. vocally). The perception-production systems of SL give rise to one macroscopic modality effect, namely the simultaneous production of a significant amount of contrastive phonemic (and morphemic) material (Vermeerbergen et al. 2007). Phonological contrast is accounted in SL phonology via feature geometry (Brentari 1998, Sandler and Lillo-Martin 2006, van der Kooij 2002). These models introduce three major classes of phonemes of primitives (*handshape*, *place of articulation* and *movement*), and derive a forth, *orientation*, as the result of the interaction between handshape and place of articulation. In other words, hand-*orientation* is not computed in absolute terms with respect to the signer's body serving as a landmark, but it is defined in relative terms. *Absolute orientation* is typically left as a phonetic implementation or as a lexical specification in iconically motivated signs (van der Kooij 2002).

Problematic Data. A source of problematic data has to do with the inventory of features that is crosslinguistically attested and necessary to derive *relative orientation*. A second source are those signs that do not meet descriptive adequacy at the *phonological* level, if *relative orientation* only is considered.
The first case is illustrated by the minimal pair EGG / SHIT in LSF, fig. (1a) and (1b). The two signs are identical except for *relative orientation*. Specifically, the radial part of the non-dominant hand is involved for EGG, while the web between the two selected fingers is involved in SHIT. The current set of features for orientation cannot capture this contrast. Orientation for EGG is derived with a $[+radial]$ feature on both hands. Orientation for SHIT has the same feature for the dominant hand, $[+radial]$, but no specification can capture the relevant part of the non-dominant hand. Indeed, Brentari (1998) explicitly exclude $[\pm web(ing)]$ from the set of active features. According to her, apparent contrasts for ASL can be derived as cases of $[+radial]$. The other models implicitly assume Brentari's proposal.

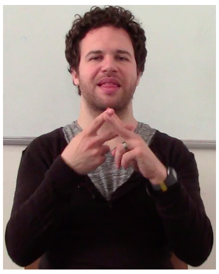
(1) a. EGG: final frame

c. ALL: initial frame



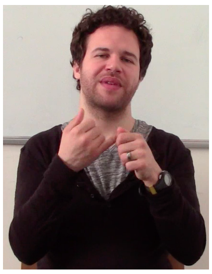
e. STRING BEAN: initial frame

b. SHIT: final frame



d. ALL: final frame

f. STRING BEAN: final frame



The second case is provided signs like ALL and STRING BEAN, fig. (1c)-(1d) and fig. (1e)-(1f). Both signs involve orientation change and are problematic for different reasons. *Relative orientation* for ALL is captured by specifying $[+tip]$ & $[+palm]$. However, this specification does not capture the fact that *relative orientation* is kept constant throughout articulation while the sign involve orientation change. Current models would predict surface forms that are near minimal pairs FLIP-PAGE and LATE (not illustrated here).

Both relative and absolute *orientations* are problematic for STRING BEAN. One could try to derive *relative orientation* of the initial frame (fig. 1e) with $[+tip]$ & $[+fingerfront]$. However, this would leave unaccounted *relative orientation* of the final frame (fig. 1f), which cannot be reasonably derived via redundancy. *Absolute orientation* is also problematic, as the orientation change leaves the main features unaffected ($[+tip]$ always faces $[+fingerfront]$). Other problems that are raised by signs like STRING BEAN will be illustrated during the presentation.

Analysis. Fixing the first problematic case does not introduce major consequences for current frameworks. These are normally modeled after one particular sign language (ASL, NGT, Israeli SL) and tacitly extended to others. It is expected that the feature inventory may not capture typological variation. The EGG / SHIT contrast in LSF simply shows the need to add a feature to the inventory of possible contrastive features in sign language. Descriptive adequacy is met by introducing a $[\pm web]$ orientation feature in the pool (see Liddell & Johnston 1989), while explanatory adequacy is met by allowing individual languages to select that feature as phonologically contrastive. Alternative solutions, like positing a $[\pm webing]$ movement feature (cf. Stokoe 1960) will be discarded during the presentation.

The second case is more problematic: the effects of orientation change cannot be captured by *relative orientation*. These are far from being just phonetic adjustments, and iconic motivations cannot be argued for. Indeed, ALL is not an iconic sign, while STRING BEAN has an iconic handshape which is not affected by orientation change. In order to account for this second set of data, we propose to introduce in the feature geometry system a second (recursive) layer for orientation specification. This would be a "secondary" plane of articulation. For sake of illustration, we take it to be the signer's body in the case of ALL and STRING BEAN. The configuration $[+tip]$ / $[+torso]$ would capture the final frame in ALL (fig. 1d), while $[+tip]$ / $[+palm]$ captures *relative orientation*. Similarly, $[+palm]$ / $[+torso]$ would capture *absolute orientation*, while $[+tip]$ / $[+fingerfront]$ captures *relative orientation*. In both cases, the initial status of absolute orientation can be redundantly recovered by movement features.

At the global level, our account introduces a major innovation in sign representation as it call for *absolute orientation*. However, it does so by minimally modifying current frameworks. Indeed, the requirement to obtain absolute orientation is to have a "secondary" plane of articulation. Notice that by specifying only one pair of features for absolute orientation leaves ample margins for phonetic adjustment, so that the flexibility required by the cases discussed in van der Kooij (2002) are still accounted for.

Conclusions. Theoretical models are extremely important to capture linguistic generalizations. However, blind extension from a language to another may lead to empirical inaccuracies. We showed that the inventory of active features is not fixed in SL. Exotic features (e.g., $[\pm web]$) may be active in creating minimal pairs in some SLs but not in others.

Reasons of elegance and economy have led researchers to eliminate *absolute orientation* from SL description. Data from LSF showed that this move is premature and that both *absolute* and *relative orientations* are needed to meet descriptive and explanatory adequacy.

References

Brentari, Diane. 1998. *A prosodic model of sign language phonology*. MITPress **Liddell & Johnson**. 1989. American Sign Language: The Phonological Base. *Sign Language Studies* 1064(1). 195277. **Sandler & Lillo-Martin**. 2006. *Sign Language and Linguistic Universals*. CUP. **Stokoe**. 2005. Sign language structure: An outline of the visual communication systems of the American Deaf. *Journal of Deaf Studies and Deaf Education* 10(1). 337. **van der Kooij**. 2002. *Phonological Categories in Sign Language of the Netherlands: The Role of Phonetic Implementation and Iconicity*. UvA. **Vermeerbergen et al.** 2007. Simultaneity in Signed languages. JBPco.

The significance of frequency effects in stylisation patterns. LOT unrounding and PRICE monophthongisation in popular music singing accent

Monika Konert-Panek
University of Warsaw

Singing accent stylisation is a phenomenon that has been extensively studied, in particular with regard to Americanisation it frequently involves (Trudgill 1983, Simpson 1999, Beal 2009, Gibson and Bell 2012 among others). Depending on the analytical framework, the concepts of identity, reference style or default accent have been perceived as main explanatory forces. Yet, some questions regarding the mechanisms of this type of stylisation call for further research, among them – the distinct behaviour of different words exhibiting a given feature, the phenomenon that may be discussed in the light of lexical frequency effects.

This paper focuses on two processes typical of singing accent Americanisation: LOT unrounding and PRICE monophthongisation, with its aim being to examine the potential significance of frequency effects from the perspective of the usage-based paradigm (Bybee 2001), represented as an exemplar model (Johnson 1997, Pierrehumbert 2001), in which sociophonetic variation plays a pivotal role. Frequent words – apart from being processed faster and articulated more easily (e.g. Bybee 2002, Shockey 2003, Erker and Guy 2012) – are said to be the leaders of various sound changes, mainly reductive ones (e.g. Hooper (Bybee) 1976, Hay, Jannedy and Mendoza-Denton 1999, Bybee 2000).

In order to assess whether frequent words prove to be the best carriers of Americanised singing style, the singing accents of two British vocalists – Amy Winehouse and Adele Laurie Blue Adkins – were analysed with respect to the selected processes. The analysis was based on five studio albums released by the artists: *Frank* and *Back to Black* by Amy Winehouse and *19*, *21* and *25* by Adele, as well as selected interviews to compare their singing and speaking styles. The data was extracted from all the albums to compile four corpora including 278 and 194 LOT tokens, and 1334 and 785 PRICE tokens in the singing styles of Adele and Winehouse, respectively. For reference, spoken styles of both vocalists were analysed (altogether 17 LOT tokens and 241 PRICE tokens). Both auditory and acoustic methods were used for the analysis of the audio material. PRAAT (Boersma and Weenink 2016) was used to provide acoustic verification of the auditory analysis whenever isolated vocal tracks were available (35 LOT tokens and 106 PRICE tokens). All the spoken tokens were analysed acoustically. PRAAT script was used for both the LOT and PRICE vowels (the midpoint method and proportional distance method, respectively). Frequency was measured locally, i.e. the word frequency of the sample itself was used, following Hay at al. (1999) and Erker and Guy (2012). All LOT and PRICE items in the corpora were identified as frequent (occurring 5 or more times) or infrequent (occurring fewer than 5 times). The statistical significance of the obtained results regarding lexical frequency effects was verified by means of a chi-square test with Yates' correction.

Acoustic analysis of the LOT vowel in the speaking mode confirms auditory impressions regarding the lack of Americanisation. However, the PRICE diphthong was monophthongised in 16% of cases in Adele's speaking style and in 40% of cases in Winehouse's speaking style, which may be suggestive of casual speech reduction process in the latter case. With regard to the singing mode, in three out of four analysed corpora, the percentage of frequent words undergoing the change was higher compared with infrequent ones (PRICE monophthongisation: 43% vs. 4% in Adele's style, 45% vs. 14% in Winehouse's style. LOT unrounding: 66% vs. 46% in Winehouse's style) and in one case infrequent words favour the process (LOT unrounding in Adele's style: 49% vs. 38%). The results in the three cases where lexical frequency favours the process are statistically significant and in one opposite case the result proves statistically insignificant, which suggests that word frequency may affect singing style variation.

References

[1] Beal, J. 2009. “You're not from New York City, you're from Rotherham”: Dialect and identity in British indie music. *Journal of English Linguistics* 37(3), 223-240.
[2] Boersma, P., & Weenink, D. 2016. *Praat: doing phonetics by computer*. Computer program.
[3] Bybee, J. 2000. The phonology of the lexicon: evidence from lexical diffusion. In Barlow, M. & Kemmer, M. (Eds.), *Usage-based models of language*. Stanford, CA: CSLI, 65–85.
[4] Bybee, J. 2001. *Phonology and language use*. Cambridge: Cambridge University Press.
[5] Bybee, J. 2002. Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change* 14, 261–290.
[6] Erker, D. & Guy, G. R. 2012. The role of lexical frequency in syntactic variability: Variable subject personal pronoun expression in Spanish. *Language* 88 (3), 526-557.
[7] Gibson, A. & Bell, A. 2012. Popular music singing as referee design. In Hernández-Campoy, J. M. & Cutillas-Espinosa, J. A. (Eds.), *Style-Shifting in Public. New perspectives on stylistic variation*. Amsterdam: John Benjamins Publishing Company, 139-164.
[8] Hay, J., Jannedy, S. & Mendoza-Denton, N. 1999. Oprah and /ay/: lexical frequency, referee design and style, *Proceedings of the 14th International Congress of Phonetic Studies 1999* (San Francisco, USA), 1389-1392.
[9] Hooper, J. B. 1976. Word frequency in lexical diffusion and the source of morphophonological change. In Christie, W. (Ed.), *Current Progress in Historical Linguistics*. Amsterdam: NorthHolland, 96-105.
[10] Johnson, K. 1997. Speech perception without speaker normalization: An exemplar model. In Johnson, K. & Mullennix, J. (Eds.), *Talker Variability in Speech Processing*, San Diego: Academic Press, 145-165.
[11] Pierrehumbert, J. B. 2001. Exemplar dynamics: Word frequency, lenition, and contrast. In Bybee, J. & Hopper, P. (Eds.), *Frequency effects and the emergence of lexical structure*. Amsterdam: John Benjamins Publishing Company, 137-157.
[12] Sackett, S. J. 1979. Prestige Dialect and the Pop Singer. *American Speech* 54 (3), 234-237.
[13] Shockey, L. 2003. *Sound Patterns of Spoken English*. Oxford: Blackwell Publishing.
[14] Simpson, P. 1999. Language, culture and identity. With (another) look at accents in pop and rock singing. *Multilingua* 18 (4), 343-367.
[15] Trudgill, P. 1983. Acts of Conflicting Identity. The Sociolinguistics of British Pop-Song Pronunciation. In Trudgill, P. (Ed.), *On Dialect. Social and Geographical Perspectives.* Oxford: Blackwell, 141-160.

Similarity and contrast in L1 pronunciation attrition in bilinguals

Brechtje Post and Samantha Jones
University of Cambridge

Research on L1 pronunciation attrition shows that second language learning has bidirectional effects, with bilingual productions in the L1 as well as the L2 usually falling somewhere between those of monolingual speakers (e.g. voice onset time (VOT) [1, 2]; vowel quality [3, 4]; intonation [5]). In fact, the interplay of the two languages has been found to lead to divergences from monolingual productions in particular in areas of grammar where there are similarities between L1 and L2, and where there is therefore most competition between the systems [6]. In this context, Flege's Speech Learning Model [7] provides a convincing account of bidirectional pronunciation effects, since the phonetic systems of the two languages are predicted to interact in such a way that a single merged category will be used in the production of sounds which are "similar" in both languages, in the L1 as well as the L2; cf. [8].

However, such a merger of similar sounds in the two languages could potentially compromise the maintenance of phonemic contrasts in the L1, depending on the way in which phonetic properties are exploited to signal contrast in the two languages (Spears 2006 for French nasalisation). For instance, nasalisation in French can be used phonemically, contrasting in e.g. /ɛ/ *paix* 'father' and /ɛ̃/ *pin* 'pine', but it also occurs as a coarticulatory allophonic realisation in the context of a nasal consonant, e.g. [ɛ̃] *peine* 'effort/punishment'. In English, by contrast, nasalisation is only coarticulatory, e.g. /ɛ/ *pet* vs. [ɛ̃] *pen*. This implies that if L1 French learners of English were to develop hybrid 'intermediate' realisations across the board, their L1 nasal contrast on vowels like /ɛ/ - /ɛ̃/ could be compromised.

In this paper, we investigate to what extent contrastivity might interact with similarity in the L1 pronunciation of late bilinguals. We analysed nasalisation in tokens of French /ɑ/, /ɛ/ and /u/ in oral and nasal contexts (CVC vs. CVN), contrasting them with nasal vowels /ɑ̃/ and /ɛ̃/, as well as "similar" English vowels /a/, /ɛ/ and /u/ which were also elicited in nasal and oral contexts. A VOT condition was included as a control to verify that our bilinguals were indeed comparable to those reported in the literature [1]. The productions of 5 French late bilingual speakers of English who had all lived in the UK for at least 10 years were compared with those of two monolingual control groups.

The results show that the bilinguals' productions do indeed fall between those of monolinguals (Figure 1) [1, 2, 3, 4], but also that bidirectional effects of a merger driven by similarity are overridden when an L1 contrast is at risk: Contrastive L1 phones which have a corresponding phone in L2 do not show signs of significant merger if contrast needs to be maintained (Figure 2: left and middle bar vs. right bar). We conclude that contrastiveness competes with similarity to determine category membership in the bilingual's shared phonetic space in unexpected ways.

This study confirms that the L1 continues to change under the influence of an L2 during adulthood. Our findings also show that, although similar phones can indeed be merged in the bilingual speaker's system - as predicted by the Speech Learning Model - mergers between similar phones are blocked when they threaten to undermine contrastivity in the native system. Moreover, even contrastive L1 phones which do not exist in the L2 can exhibit signs of change after L2 acquisition, showing L2 effects that are not directly motivated by similarity or contrastiveness. This suggests that bilingual speech development is not only multilayered and systematic, but that the interactions between the L1 and L2 in the developmental process are truly systemic in nature, affecting comparable elements throughout the shared bilingual system in a similar way, while both similarity and contrastiveness between the elements in the L1 and L2 delimit convergence and divergence between the two, cf. [9].

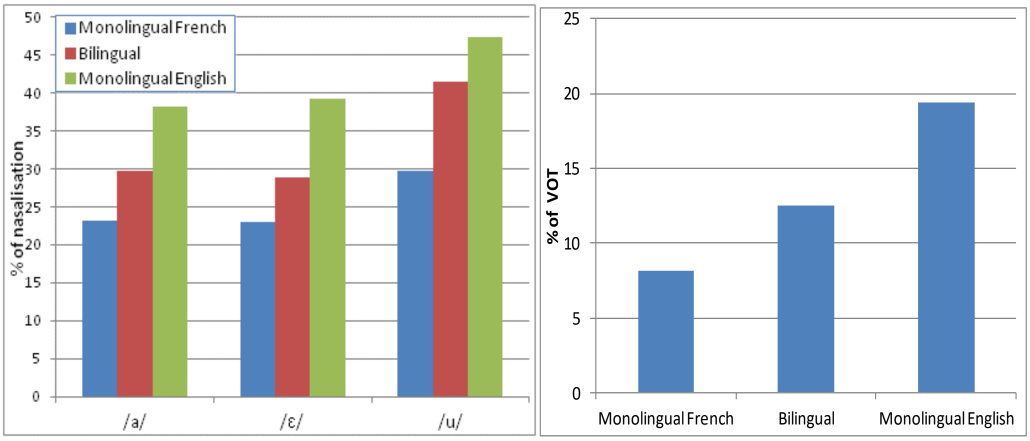


Figure 1. *Mean coarticulatory nasalisation as percentage of vowel duration (left) and VOT (/t/ + /u/) as percentage of syllable duration (right) for mono- and bilinguals.*

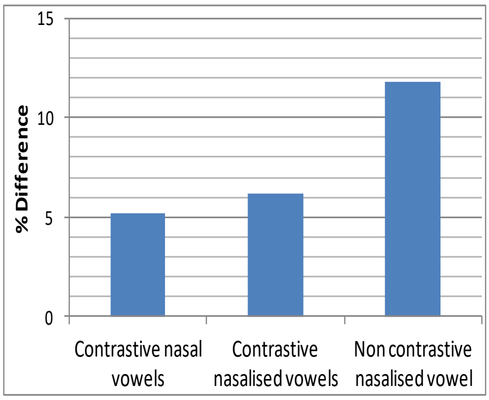


Figure 2. *Difference (%) between bilingual and monolingual speakers for contrastive nasal* /ã, ɛ̃/, *nasalised allophonic* [ã, ɛ̃] *(is contrastive in French) & nasalised allophonic* [ũ].

[1] Flege, J., E. (1987). The Production of "new" and "similar" phones in a foreign language: evidence for the effect of equivalence classification. *Journal of Phonetics* 15: 47-65.

[2] Major, R. (1992). Losing English as a first language. *The Modern Language Journal* 76 (2): 190–208.

[3] Jiang, H. (2008). *Effect of L2 phonetic learning on L1 vowels*. Doctoral thesis Simon Fraser University, Canada.

[4] Chang, C. B. (2012). Rapid and multifaceted effects of second-language learning on first-language speech production. Journal of Phonetics, 40(2), 249-268.

[5] Mennen, I. (2004). Bi-directional interference in the intonation of Dutch speakers of Greek. *Journal of Phonetics*, 32, 543-563.

[6] Schmid, M.S. (2010). Languages at play: The relevance of L1 attrition to the study of bilingualism. *Bilingualism: Language and Cognition* 13(1), 1–7.

[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. 233–272). Baltimore, MD: York Press.

[8] Lord, G. (2008). Second language acquisition and first language phonological modification. In: J. B. de Garavito & E. Valenzuela (Eds.), Selected proceedings of the 10th Hispanic linguistics symposium (pp. 184–193). Somerville, MA: Cascadilla Proceedings Project.

[9] De Bot, K., Lowie, W., & Verspoor, M. (2007). A Dynamic Systems Theory approach to second language acquisition. *Bilingualism: Language and Cognition*, 10, 7–21.

Beat Gestures as Prosodic Domain markers in French: A case study

Patrick Louis Rohrer[1,2] & Elisabeth Delais-Roussarie[2]

[1]Department of Translation and Language Sciences, Universitat Pompeu Fabra, Barcelona, Catalonia
[2]Laboratoire de Linguistique de Nantes (LLING) – UMR 6310, Université de Nantes, Nantes, France

It is widely held that co-speech gestures are produced in a coordinated fashion with prosodic prominence [1, 2, among many others]. Studies have shown that gesture apexes tend to be temporally executed in conjunction with pitch accentuation [3, 4, and 5, among many others]. Fewer studies have looked at the temporal execution of these gestures in languages such as French, where pitch accentuation operates on the domain of prosodic phrases. A main pitch accent (T*) obligatorily marks the right edge of the smallest prosodic phrase (also known as the accentual phrase, or AP, see [6]), and a pitch rise may or may not occur on the initial syllables of the AP, marking the left edge. In this sense, these accents are said to have a demarcative function (among other pragmatic functions such as rhythmic marking or emphasis). Prosodic and gestural analyses were carried out on an 18-minute, academic-style discourse, with the goal of exploring the temporal relationship between gesture production and prosodic structure in French. Specifically we ask if the apex of beat gestures tends to coincide with pitch accented syllables. In cases where beat apexes do not coincide with a pitch accented syllables, do beat gestures continue to mark phrase-initial or –final positions independent of pitch accentuation?

The discourse for study comes from a Ted Talk presentation by [7]. This speaker was chosen for his extensive use of gestures, as well as the video editing of his presentation, which kept the amount of time where the speaker was not visible to a minimum. The speaker's gestures were annotated by the first author in ELAN and included annotations for gesture phrasing (preparation/stroke/retraction) and apex. The coding was carried out without accompanying audio in order to avoid influence from audio in determining apex placement. The apex was determined following [8: 190], where the endpoints of unidirectional strokes, or moments of change in velocity or direction of multi-directional strokes were considered as apexes. Gestures were then reassessed with audio in order to determine their referentiality and each gesture was classified into one of McNeill's four gesture dimensions [2]. Regarding prosodic annotations, the F_ToBI system [6] was used to determine prosodic phrasing and pitch accentuation. Further, as gesture apexes that occur up to 200ms before a stressed syllable may perceptively align [9], any syllable that was within 150ms of a pitch-accented syllable was perceptively double checked and marked as perceptively aligning or not.

Gesture coding resulted in a total of 779 apexes. Of these, 670 apexes were classified as coming from non-referential beat gestures. Twenty-five apexes were excluded from further analyses due to their occurrence outside of speech, or they occurred with another beat apex on the same syllable. Beat apexes occurred perceptively during pitch accented syllables 72.09% of the time. Beat gesture apexes occurred during non-accented syllables 27.91% percent of the time. Of these, 72.23% still occurred on AP-initial syllables.

These results suggest that beat gestures are often temporally correlated with pitch accentuation (see Figure 1), but when pitch accentuation is not present, beat gestures may continue to mark prosodic phrasing independent of pitch accentuation (see Figure 2). These findings suggest that the relationship between gesture production and pitch accentuation may not be as straight-forward as previously reported, and a language's prosodic structure may influence gesture production. Further research is needed to disambiguate other factors that may influence gesture placement independent of pitch accentuation, such as the role of prosodic lengthening or rhythmic marking.

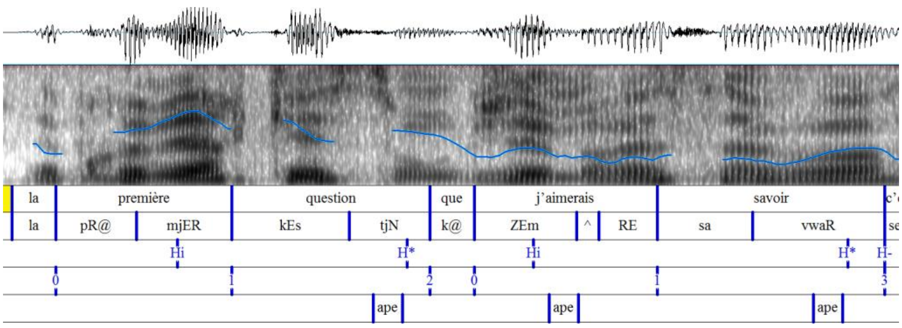


Figure 1. *F_ToBI and gesture annotations of a 2-AP phrase "la première question que j'aimerais savoir" [The first question I would like to know] where the right edge of the first AP and both the right and left edge of the second AP are marked both prosodically and gesturally*

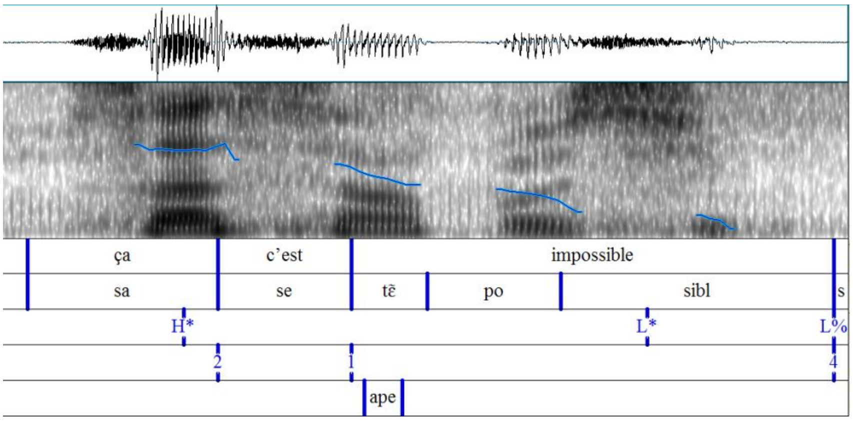


Figure 2. *F_ToBI and gesture annotations of a 2-AP phrase "Ça, c'est impossible" [in English "This is impossible"] where the second AP's left edge is marked by a beat gesture but not by prosody.*

[1] Kendon, A. 1980. Gesticulation and speech: Two aspects of the process of utterance. In Key, M.R. (Ed.), *The relationship of verbal and nonverbal communication.* The Hague: Mouton.

[2] McNeill, D. 1992. *Hand and mind: What gestures reveal about thought*. Chicago: Chicago University Press.

[3] Loehr, D. 2012. Temporal, structural, and pragmatic synchrony between intonation and gesture. *Laboratory Phonology* 3(1), 71-89.

[4] Shattuck-Hufnagel, S. & Ren, A. 2018 in press. The Prosodic Characteristics of Non-referential Co-speech Gestures in a Sample of Academic-Lecture-Style Speech. *Frontiers in Psychology*.

[5] Esteve-Gibert, N. & Prieto, P. 2013. Prosodic structure shapes the temporal realization of intonation and manual gesture movements. *Journal of Speech, Language, and Hearing Research*, 56(3), 850-864.

[6] Delais-Roussarie, E., Post, B., Avanzi, M., Buthke, C., Di Cristo, A., Feldhausen, I., Jun, S.A., Martin, P., Meisenburg, T., Rialland, A., Sichel-Bazin, R., & Yoo, H.Y. 2015. Intonational phonology of French: Developing a ToBI system for French. In Frota, S. & Prieto, P. (Eds.), *Intonation in Romance*, Oxford: Oxford University Press.

[7] Laroche, D. 2016. "Et si votre rêve devenait possible…" [online] Tedx talk available at https://www.youtube.com/watch?v=fbGfe78_2jc

[8] Loehr, D. 2007. Aspects of rhythm and gesture in speech. *Gesture*, 7(2), 179-214.

[9] Leonard, T. & Cummins, F. 2011. The temporal relation between beat gestures and speech. Language and Cognitive Processes. 26(10), 1457–1471.

Signal- and expectation-based processing of pitch accent types

Christine T. Röhr[1], Stefan Baumann[1], Martine Grice[1], Petra B. Schumacher[2]
[1]*IfL Phonetik, University of Cologne*, [2]*IdSL I, University of Cologne*

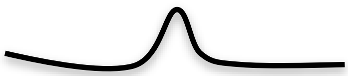
In many languages speakers employ prosody to highlight new or surprising information, making it more prominent, while they attenuate given or expected information, making it less prominent. For German it has been shown that pitch accent types differ in their degree of perceived prominence [1] and play an important role in the marking of information status (e.g. [2]). The pitch movement leading towards the target on the accented syllable ('onglide' [3]) is the major tonal cue for prominence, rising onglides being perceptually more prominent than falling ones [1]. A previous ERP study [4] found that different pitch accent types were mapped onto information status in real-time and suggested that the processing of prosodic cues used by the speaker to (re)orient the listener's attention interacts with expectations evoked by the pragmatic context. The present ERP study further investigates this interaction between signal- and expectation-driven processing, involving pragmatic contexts that generate an expectation for a particular prosodic form (L+H* vs. H+L* accent) and thus lead to differential neural correlates of expected or mismatching informational processing.

A set-up with two discrete pre-contexts for each test sentence (60 items) was designed to trigger expectations about appropriate upcoming prosody. For example, the pre-context in (1a) builds up an expectation for new, exciting information, guiding the attention of the listener to an event that should be highlighted by prosodic prominence. The pre-context in (2a) establishes that nothing new or unusual is going to follow and builds up an expectation for neutral information. Test sentence (1b), with a prominent accent (L+H*, rising onglide) appropriate for new, exciting information, matches context (1a) but not context (2a). Conversely, test sentence (2b), with a less prominent accent (H+L*, falling onglide), typically found on contextually derivable information, matches context (2a) but not (1a).

We hypothesize that mismatches (incongruent combinations of prosody and context expectation) will result in an expectation-based error at an early discourse processing stage which has been found to engender a negative ERP deflection (N400) (e.g. [5, 6]). Furthermore, signal-driven attention orienting (involving prominence-lending cues) requires listeners to update their mental model at a later discourse processing stage. Such mechanisms may engender a late positive ERP deflection, e.g. when expressing topic shift or for prosodically marked new information and focused constituents (e.g. [6, 7]). Given that rising onglides are perceptually more prominent than falling onglides, we predict that the former will trigger more attentional orienting. However, previous results were mixed, since unexpected accents, too, may cause a late positivity (in addition to an N400) (e.g. [4, 6]).

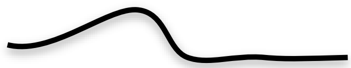
Twenty-four monolingual native speakers of German (19f, 5m) participated in the current study. They listened to context and test sentence (240 test items + 120 filler items) and performed a word recognition task. We calculated regression-based ERPs [8] and computed linear regression models (factors: rising/falling onglide and exciting/neutral context; continuous covariates: pitch and intensity). Results (see Fig.1) indicate for both accent types a significantly more negative deflection (N400) following the mismatching context. Furthermore, both accent types elicit a late positivity following the exciting context. This may be tied to a mechanism of attention orienting triggered by a highlighting effect of the exciting context. Rising compared to falling onglides also show an early positivity. Hence, the results provide evidence for discrete neurophysiological correlates of prosodic and content-related prominence: Unexpected accents yield an expectation error (N400) while a prominent rising accent consumes attentional resources and engenders updating processes (early positivity) which is also observed for content-related highlighting (late positivity) independent of the prosodic realization of the target word.

(1) (a) *Rate mal, was uns heute passiert ist!*
‘Guess what happened to us today!’
exciting context



(b) *Wir haben Mi**LE**na getroffen.*
‘We met Milena.’
L+H* (rising onglide)

(2) (a) *Heute ist nichts Besonderes passiert.*
‘Today, nothing special happened.’
neutral context



(b) *Wir haben Mi**LE**na getroffen.*
‘We met Milena.’
H+L* (falling onglide)

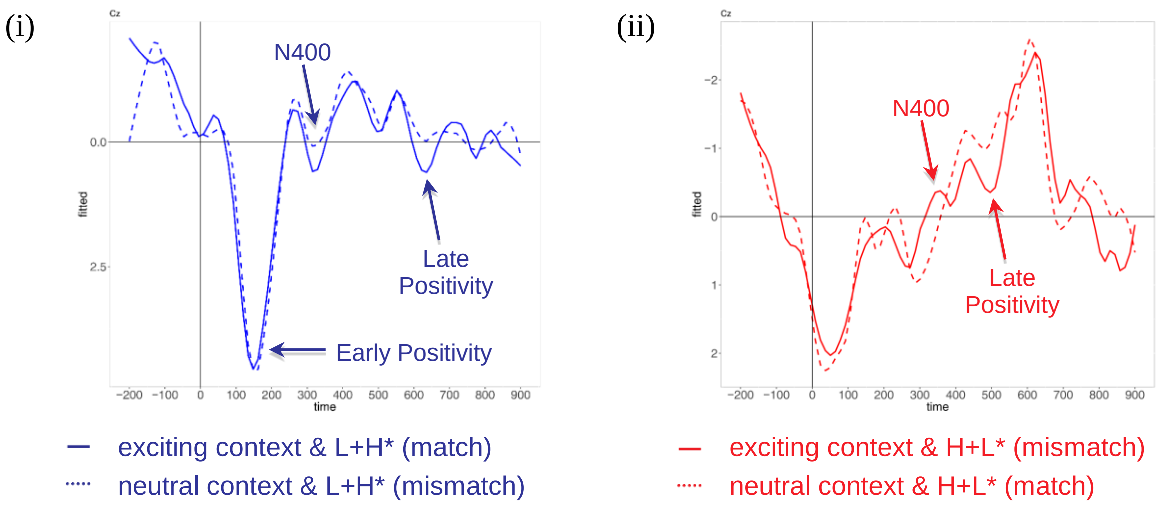


Figure 1. *Grand average rERPs at a selected electrode (CZ) from 200 ms before until 900 ms after the critical word (= name/object; onset at vertical bar). (i) shows the rERPs for L+H* accents (rising onglide), (ii) for H+L* accents (falling onglide). Exciting (solid lines) and neutral contexts (dotted lines) yield different (mis)matches depending on target prosody.*

[1] Baumann, S. & Röhr, C.T. 2015. The Perceptual Prominence of Pitch Accent Types in German, *Proc*. *ICPhS XVIII*, Glasgow, UK, vol. 298, 1-5.

[2] Röhr, C.T. & Baumann, S. 2010. Prosodic Marking of Information Status in German, *Proc. 5th International Conference on Speech Prosody*, Chicago, USA, vol. 100019, 1-4.

[3] Ritter, S. & Grice, M. 2015. The Role of Tonal Onglides in German Nuclear Pitch Accents. *Language and Speech*, 58(1), 114-128.

[4] Schumacher, P.B. & Baumann, S. 2010. Pitch Accent Type Affects the N400 During Referential Processing. *NeuroReport*, 21(9), 618–622.

[5] Kutas, M. & Federmeier, K.D. 2000. Electrophysiology Reveals Semantic Memory Use in Language Comprehension. *Trends in Cognitive Sciences* 4(12), 463-470.

[6] Toepel, U., Pannekamp, A. & Alter, K. 2007. Catching the News: Processing strategies in listening to dialogs as measured by ERPs. *Behavioral and Brain Functions* 3, 53.

[7] Hruska, C. & Alter, K. 2004. Prosody in dialogues and single sentences: How prosody can influence speech perception. In Steube, A. (Ed.), *Information Structure. Theoretical and Empirical Aspects*. Berlin, New York: De Gruyter, 211-226.

[8] Smith N.J. & Kutas, M. 2015a/b. Regression-based estimation of ERP waveforms. *Psychophysiology* 52(2), 157-181.

Development of an L2 Japanese speech corpus
for the comparison of prosody across diverse L1 groups

Motoko Ueyama[1] Ryoko Hahashi[2]
Aaron Lee Albini[2]

[1]*University of Bologna,* [2]*Kobe University*

Prosody is a significant contributing factor in how native speakers rate the pronunciation of second language (L2) learners [1]. Compared to individual segments, learners' divergences from target-like prosody have been reported to have a stronger effect on native speakers' evaluations [2] and lead to more serious communication difficulties [3]. In the case of L2 Japanese, prosody is known to be one of the most challenging aspects of Japanese pronunciation for learners to acquire [4]. To better understand which specific aspects of prosody pose the greatest problems (and therefore are of the greatest pedagogical importance), there is a need for large-scale databases of experimentally elicited samples of learner speech [5]. Unfortunately, such resources are still scarce. While a handful of such corpora do exist for L2 Japanese (see [6]), none of them explicitly target a wide range of prosodic phenomena while simultaneously sampling from a diverse set of native-language (L1) groups.

The present study reports on an on-going corpus construction project to fill this gap. At present, the corpus consists of Japanese learners mainly at the upper elementary level, representing four different L1 backgrounds: Chinese (N=29), Korean (N=23), Italian (N=16), and Russian (N=15). In addition, a sample of 9 native speakers of Tokyo Japanese is also included as a baseline. The stimulus materials were designed to examine five aspects of Japanese sentence-level prosody: (1) final rise intonation in interrogatives, (2) focal prominence, (3) downtrends (especially downstep), (4) the disambiguation of syntactically ambiguous sentences, and (5) the prosodic realization of emotions and other paralinguistic information. To serve as a point of reference for comparing the acquisition of prosody vs. segmental phenomena, a sixth set of materials was also included testing vowel devoicing. Data is being collected using the Online Voice Recorder tool referenced in [7]. Upon completion of the project, the corpus will be made publicly available online.

A preliminary analysis comparing the L1 Italian data to the native Tokyo Japanese baseline revealed learners to diverge from native norms in the following ways:

1. **Interrogatives**: Learners often inserted a superfluous pitch accent on the penultimate syllable of lexically unaccented words, but only when produced in interrogatives (i.e., not in the parallel declaratives).
2. **Prominence**: Many learners (A) produced sentences with no prominence at all, (B) placed prominence on an unfocused word, or (C) realized prominence by inserting a superfluous pitch accent on a lexically unaccented word.
3. **Downtrends**: Several learners (especially less fluent ones) failed to produce the native-like stairstepping sequence of downsteps.
4. **Disambiguation**: Many learners' strategies involved inserting a long pause and/or dividing the sentence into multiple IPs (in contrast to the native speakers, who generally phrased each sentence as a single IP).
5. **Paralinguistics**: While learners could produce *joy* and *surprise* in native-like ways, this was not true for *anger*, *sadness*, and *seriousness*.
6. **Devoicing**: None of the learners had difficulty devoicing the relevant vowels.

The unique design of the corpus makes two kinds of comparison possible. First, through analyses like those just described, each individual L1 group can be examined in terms of which of the six phenomena are most challenging. Second, the various L1 groups can be compared in terms of their performance with respect to one specific phenomenon (e.g., downstep). The results of such analyses hold great potential not only for L2 prosody research but also for the future development of tools and materials for teaching L2 Japanese.

References

[1] Lengeris, A. (2012). Prosody and second language teaching: Lessons from L2 speech perception and production research. In J. Romero-Trillo (Ed.), *Pragmatics and Prosody in English Language Teaching* (pp. 25-40). Dordrecht, Netherlands: Springer.

[2] Sato, T. (1995). Tan'on to inristu ga nihongo onsei no hyōka ni ataeru eikyō. *Sekai no Naka no Nihongo Kyōiku*, *5*, 139-154.

[3] Celce-Murica, M., Brinton, D. & Goodwin, J. (1996). *Teaching Pronunciation: A Reference for Teachers of English to Speakers of Other Languages*. Cambridge, UK: Cambridge University Press.

[4] Toda, T. (2008). Hatsuon no tatsujin to wa dono yō na gakushūsha ka. In Toda, T., (Ed.), *Nihongo Kyōiku to Onsei* (pp. 61-80). Tokyo, Japan: Kuroshio Shuppan.

[5]Gut, U. (2014). Corpus phonology and second language acquisition. In J. Durand, Gut, U., & Kristoffersen, G. (Eds.), *The Oxford Handbook of Corpus Phonology* (pp. 286-301). Oxford, UK: Oxford University Press.

[6] Maekawa, K. (2009). Nihongo gakushūsha onsei kenkyū no kadai. *Nihongo Kyōiku*, *142*, 4-13.

[7] Yue, J., Shiozawa, F., Toyama, S., Yamauchi, Y., Ito, K., Saito, D. & Minematsu, N. (2017). Automatic scoring of shadowing speech based on DNN posteriors and their DTW. In *Proceedings of Interspeech 2017* (pp. 1422-1426). DOI: 10.21437/Interspeech.2017-728.

Producing rhythmic beat gestures while retelling a story: positive effects of a gesture-based training session on children's narrative performance

Ingrid Vilà-Giménez[1], Pilar Prieto[2,1]
[1]*Dept. of Translation and Language Sciences, Universitat Pompeu Fabra, Catalonia,*
[2]*Institució Catalana de Recerca i Estudis Avançats (ICREA), Catalonia*
ingrid.vila@upf.edu, pilar.prieto@upf.edu

Recent research with preschoolers has given evidence that observing and listening to storytellers who are also performing rhythmic beat gestures (e.g., hand gestures that typically associate with prosodically prominent positions in speech) favor story memorization and comprehension [1, 2, 3], and also lead to better subsequent narrative performance in terms of narrative structure scores in a narration task [4]. Nevertheless, to our knowledge, no study to date has addressed the possible effects of encouraging children to produce rhythmic beat gestures while retelling narratives on their performance –as opposed to merely observing them–. Our hypothesis is based on evidence that suggests that gesturing has beneficial effects on various cognitive and linguistic domains (see [5] for a review). Moreover, as beat gestures may boost narrative performance due to their role in highlighting important structural properties of language (such as focus, discourse structure and rhythm [6, 7]), we are also led to believe that they can be the root of potential beneficial effects in narrative production and development.

In this study, a total of forty-seven 5- and 6-year-old children participated in a brief between-subject training study with a pretest and an immediate posttest design (Fig. 1). Following a pretest, which consisted of recounting a cartoon they had been shown, the children watched a set of six short videotaped stories told by storytellers who performed beat gestures with prosodic prominence in target positions within the story. One half of the children were then asked to merely retell the story they had seen without any instructions regarding gesture (*beat non-encouraging condition*), while the other half were asked to retell the story they had seen and imitate the beat gestures that were used by the storyteller (*beat encouraging condition*). Both groups then performed a posttest, which was identical in procedure to the pretest. The animated cartoons used in the pretest and posttest were different from those stories of the training, but they all followed a similar narrative structure. Video recordings of the pretest and posttest narratives were then scored for narrative structure and fluency. Narrative structure was rated on a scale of 0 to 6 (6 = complete goal-based narrative). Fluency scores were exclusively based on a perceptive coding scale from 1 to 7 (1 = extremely disfluent and 7 = extremely fluent). Results for narrative structure (Fig. 2) showed a main effect of Test ($F(1, 184) = 25.194$, $p < .001$) and a significant interaction between Condition and Test ($F(1, 184) = 6.167$, $p = .014$). Fluency scores (Fig. 3) revealed a main effect of Test ($F(1, 184) = 18.277$, $p < .001$) and a significant interaction between Condition and Test ($F(1, 184) = 4.649$, $p = .032$). The findings showed that children in the group that had been encouraged to use beat gestures (*beat encouraging*) received significantly higher gains in the quality of their posttest narratives in terms of narrative structure and fluency scores compared to the group of children who had merely recounted the story without previous gesture instruction. All in all, this evidence suggests that encouraging the use of rhythmic beat gestures in children helps boost their subsequent narrative performance. This research can have an impact on our understanding of children's gesture and narrative development, as well as practical implications for teaching methodologies.

Keywords: beat gestures; prosody; narrative discourse performance; between-subject training study

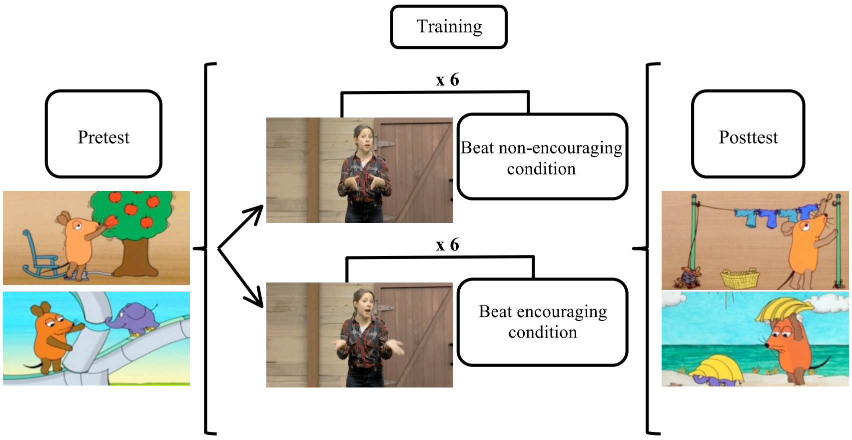
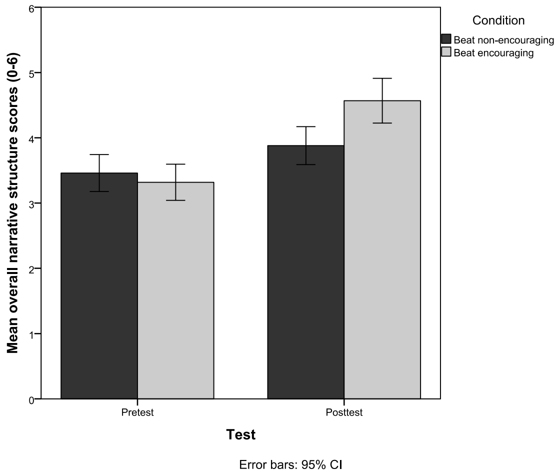


Fig. 1. *Experimental procedure*.



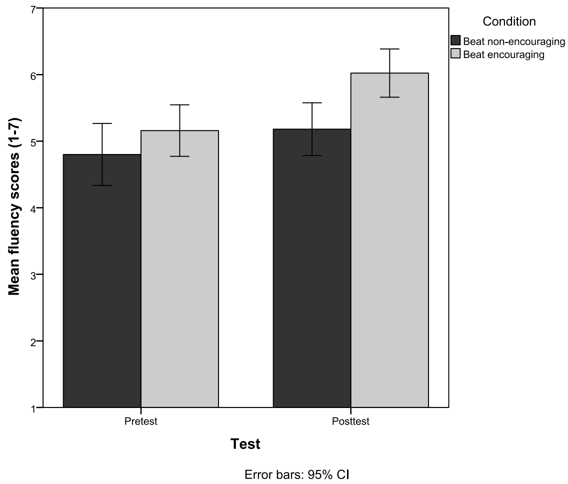


Fig. 2 & 3. *Mean overall narrative structure scores from 0 to 6 and fluency scores from 1 to 7, separated by training condition and test. Error bars represent 95% confidence intervals of the means.*

[1] AUSTIN, E., & SWELLER, N. (2014). Presentation and production: The role of gesture in spatial communication. *J Exp Child Psychol, 122*(1), 92–103.

[2] IGUALADA, A., ESTEVE-GIBERT, N., & PRIETO, P. (2017). Beat gestures improve word recall in 3- to 5-year-old children. *J of Exp Child Psychol*, *156,* 99–112.

[3] LLANES-COROMINA, J., VILÀ-GIMÉNEZ, I., KUSHCH, O., BORRÀS-COMES, J., & PRIETO, P. (2018). Beat gestures help preschoolers recall and comprehend discourse information. *J Exp Child Psychology, 172*(8)*,* 168–188.

[4] VILÀ-GIMÉNEZ, I., IGUALADA, A., & PRIETO, P. (2019). Observing storytellers who use rhythmic beat gestures improves children's narrative discourse performance. *Dev Psychol*, *55*(2)*,* 250–262.

[5] KITA, S., ALIBALI, M. W., & CHU, M. (2017). How do gestures influence thinking and speaking? The gesture-for-conceptualization hypothesis. *Psychol Rev, 124*(3), 245–266.

[6] SHATTUCK-HUFNAGEL, S., REN, A., MATHEW, M., YUEN, I., & DEMUTH, K. (2016). Non-referential gestures in adult and child speech: Are they prosodic? *Proc from the 8th International Conference on Speech Prosody* (pp. 836–839). Boston, MA.

[7] DMITROVA, D., CHU, M., WANG, L., ÖZYÜREK, A., & HAGOORT, P. (2016). Beat that word: how listeners integrate beat gesture and focus in multimodal speech discourse. *J Cogn Neurosci, 28*(9), 1255–1269.

Phonetic Convergence in Mandarin

Haoru Zhang and Chelsea Sanker
Brown University

Introduction: Convergence is the phenomenon in which people match others' behaviors in a communicative setting. Although a wide range of phonetic *variables* have been identified that exhibit convergence, e.g. F0 (Babel & Bulatov, 2012), formants, and vowel duration (Sonderegger, Bane & Graff, 2017), most previous convergence studies have examined a small number of *languages*, with the vast majority using English.

Some work suggests that phonological contrastiveness limits converge (Podlipský & Šimáčková, 2015) or, conversely, increases convergence based on increasing attention to particular details (Mitterer & Ernestus, 2008). However, comparisons are often within a single language, so differences between measures may be due to the phonological system or may reflect phonologically independent characteristics of each measure. Data from tonal languages, in comparison to previous work on non-tonal languages, can provide insight into the role of phonology in phonetic convergence. The current study expands on previous work by examining controlled experimental data from a widely spoken tonal language—Mandarin.

Methods: 42 Mandarin speakers (age 18-30, 19 female) were recruited in Beijing, China and assigned in two experimental paradigm groups (shadowing and exposure). Participants heard 58 monosyllabic CV Mandarin words spoken by a native Mandarin speaker, split into three blocks. For the shadowing group, participants repeated after each word; for the exposure group, participants listened to a block of words and then read the same words. All participants read the target words aloud before the listening task, produced them three times during the task, and read them again afterwards. F0 maximum, F1, F2, and vowel duration were measured, for all four tones and seven vowels. Data was modeled with linear mixed-effects regression.

Results: The results have been pooled across experimental groups, as there was no effect of task type. Convergence was found in Mandarin similar to that previously observed in other languages; the model talker's productions were predictive of participants' post-task speech. In addition, phonologically contrastive categories shaped convergence effects for F0 maximum, F1 and F2; these measures exhibited both overall convergence and category-specific convergence.

Speakers exhibited significant convergence in F0 maximum (shown in Figure 1), comparable to F0 convergence in English, but additional convergence effects were sensitive to particular tone categories. Speakers also exhibited significant convergence in vowel formants, sensitive to vowel category; F1 is shown in Figure 2 (F2 exhibited similar results, omitted due to space). The main effect suggests convergence shifting the overall vowel space, while vowel-specific effects indicate that category membership shapes convergence. These results parallel previous results in English. Finally, there was also significant convergence in vowel duration, as shown in Table 3. In comparison to formants and F1, it was not sensitive to phonological categories, either tone categories, as illustrated in Figure 3, or vowel categories.

Conclusions: The results offer support for phonetic variation shaped by phonological contrasts, as phonological categories form salient domains within which convergence occurs. There are also broad patterns of convergence within a phonetic characteristic, but convergence can vary within the specific categories that contrast in that measure. While Mandarin speakers, like English speakers, converge broadly to speaking higher or lower in their F0 range, this shift is not reflected equally across tone categories, which may suggest different salience of F0 by category, producing different degrees of attention and subsequent convergence.

	Estimate	Std. Error	df	t value	p value
(Intercept)	2.99	0.32	79.3	9.47	< 0.0001***
Pretest	0.24	0.023	1783	10.97	< 0.0001***
Model Talker (Convergence)	0.17	0.055	64.8	3.19	0.0022**
Tone - Set 1	0.79	0.34	57.7	2.32	0.024*
Tone - Set 2	-0.92	0.63	57.2	-1.47	0.15
Tone - Set 3	-0.06	0.37	69.5	-0.17	0.86
Model * Tone Set 1	-0.14	0.062	57.7	-2.29	0.026*
Model * Tone Set 2	0.16	0.12	57.3	1.42	0.16
Model * Tone Set 3	0.002	0.067	69.4	0.03	0.98

Figure 1. Regression model for post-task F0-maximum. Intercept : Tone = Grand mean of tones.

	Estimate	Std. Error	df	t value	p value
(Intercept)	0.038	0.039	376.7	0.97	0.33
Pretest	0.25	0.026	1833	9.8	< 0.0001***
Model Talker (Convergence)	0.45	0.097	34	4.7	< 0.0001***
Vowel - Set 1	0.88	0.15	16.6	5.7	< 0.0001***
Vowel - Set 2	0.048	0.049	169.7	0.97	0.33
Vowel - Set 3	-0.37	0.084	22.1	-4.45	< 0.0001***
Vowel - Set 4	-0.015	0.079	122.3	-0.19	0.85
Vowel - Set 5	0.37	0.15	563.6	2.41	0.016*
Vowel - Set 6	-0.29	0.061	138.9	-4.78	< 0.0001***
Model * Vowel Set 1	-0.34	0.10	733.2	-3.25	0.0012**
Model * Vowel Set 2	-0.11	0.11	547.8	-0.95	0.34
Model * Vowel Set 3	-0.11	0.10	1029	-1.12	0.26
Model * Vowel Set 4	-0.04	0.25	954.3	-0.16	0.87
Model * Vowel Set 5	1.35	0.41	1330	3.31	< 0.0001***
Model * Vowel Set 6	-0.33	0.12	189.2	-2.81	0.0054**

Figure 2. Regression model for post-task F1. Intercept : Vowel = Grand mean of vowels.

	Estimate	Std. Error	df	t value	p value
(Intercept)	0.13	0.026	57.2	4.93	< 0.0001***
Pretest	0.22	0.025	1623	8.98	< 0.0001***
Model Talker (Convergence)	0.35	0.088	50.7	3.97	< 0.0001***
Tone - Set 1	-0.046	0.055	44.7	-0.84	0.4
Tone - Set 2	-0.032	0.046	44.7	-0.69	0.49
Tone - Set 3	0.06	0.031	47.9	1.97	0.055 .
Model * Tone Set 1	0.13	0.185	44.3	0.7	0.49
Model * Tone Set 2	0.24	0.167	44.3	1.44	0.16
Model * Tone Set 3	-0.13	0.094	44.5	-1.33	0.19

Figure 3. Regression model for post-task vowel duration. Intercept : Tone = Grand mean of tones.

Babel, M., & Bulatov, D. (2012). The role of fundamental frequency in phonetic accommodation. *Language and Speech, 55*(2), 231-248.

Mitterer, H., & Ernestus, M. (2008). The link between speech perception and production is phonological and abstract: Evidence from the shadowing task. *Cognition*, *109*(1), 168-173.

Podlipský, V. J., & Šimáčková, Š. (2015). Phonetic imitation is not conditioned by preservation of phonological contrast but by perceptual salience. *International Congress of Phonetic Sciences (ICPhS) 2015*.

Sonderegger, M. Bane, M. & Graff, P. (2017). The medium-term dynamics of accents on reality television. *Language, 93(3)*, 598-640.

Oral session 3

Factors affecting perception I

The primary importance of onsets: Timing and prediction in speech segmentation

Laurence White[1], Silvia Benavides-Varela[2], Katalin Mády[3] & Sven Mattys[4]
[1]Newcastle University; [2]University of Padua;
[3]Hungarian Academy of Sciences, Budapest; [4]University of York

Speech sounds run one into the next, even across word boundaries. Whilst word recognition often suffices for speech segmentation when hearing our native language(s), structural ambiguity or lack of context may require listeners to exploit the available range of non-lexical cues to word boundaries [1]. Such cues are particularly important in difficult listening conditions and for first language (L1) or second language (L2) learners. Paradoxically for acquisition, however, some non-lexical cues – such as boundary-related phonotactic regularity – rely on language-specific knowledge, thus can only be incrementally exploited by new learners. Cross-linguistic consistency of prosodic segmentation cues may resolve this paradox: certain timing and intonational features may be language-independent and thus potentially available for use by L1 and L2 learners early in acquisition [2].

In the first series of experiments, we used an artificial language (AL) stream of trisyllabic nonwords (e.g., *nudopa*, *subako*, etc.) to compare the segmentation power of three durational cues across native English, Hungarian and Italian speakers. Word- or phrase-final vowel lengthening (here **Syl3-V**, e.g., *nudopa*) is widely observed across languages [3] and has been held to be a universal segmentation cue [4]. If so, lengthening of the first-syllable vowel in polysyllabic words (**Syl1-V**, *nudopa*) should mislead listeners into perceiving a mid-word prosodic boundary, at least in the absence of intonational cues. Finally, word-onset consonant lengthening (**Syl1-C**, *nudopa*) is seen in prosodically-diverse languages [5], although its importance for segmentation has been relatively little explored [e.g., 6 - Dutch, 7 - English] and not within a single experiment using the same materials for multiple languages.

Listeners were exposed to the AL stream for six minutes, then completing a 24-item two-alternative forced-choice task to test how their nonword recognition was affected by the three lengthening conditions (**Syl1-C**, **Syl1-V**, **Syl3-V)** relative to an evenly-timed baseline (**Flat**). (In this between-subjects design, each participant only heard one AL timing condition.) As Figure 1 shows, participants across languages showed higher nonword recognition in the **Syl1-C** condition than the **Flat** baseline (English p = .010, Hungarian p < .0001, Italian p = .012), indicating they all benefited, during AL exposure, from the segmentation support provided by word-initial consonant lengthening. This accords with the critical importance of onsets for lexical processing, as they initiate activation of word hypotheses [8]. By contrast, **Syl3-V** improved recognition over **Flat** only for English listeners (p = .002), contradicting the hypothesis that final-vowel lengthening is a universal segmentation cue, but in line with recent findings for Italian [9]. For all languages, **Syl1-V** was equivalent to or worse than **Flat**, reinforcing the conclusion that it is *localised* word-initial consonant lengthening that is key.

A second series of segmentation experiments, with native English listeners, examined how timing cues are processed, using a method designed to enable within-subject comparison of different cues. On each trial, participants heard a 12-syllable nonsense utterance (e.g., *pabikugolatudaropi*) followed by a nonword target (e.g., *golatu*) and had to decide if the target had been in the preceding utterance. We used the same duration conditions – **Flat**, **Syl1-C**, **Syl1-V**, **Syl3-V –** but varied the position of the target within the preceding carrier utterance: early, medial or late. Results reinforced the primary importance of word onsets: nonword targets were detected better than the **Flat** baseline only when onset consonants were lengthened, and then only when the **Syl1-C** targets were late in the carrier (p < .05 *vs* all other conditions). This pattern strongly suggests the use of a predictive timing mechanism [10], with listeners exploiting foregoing speech rate within utterances to gradually build up durational expectations and thus detect lengthened onset consonants as word boundary cues.

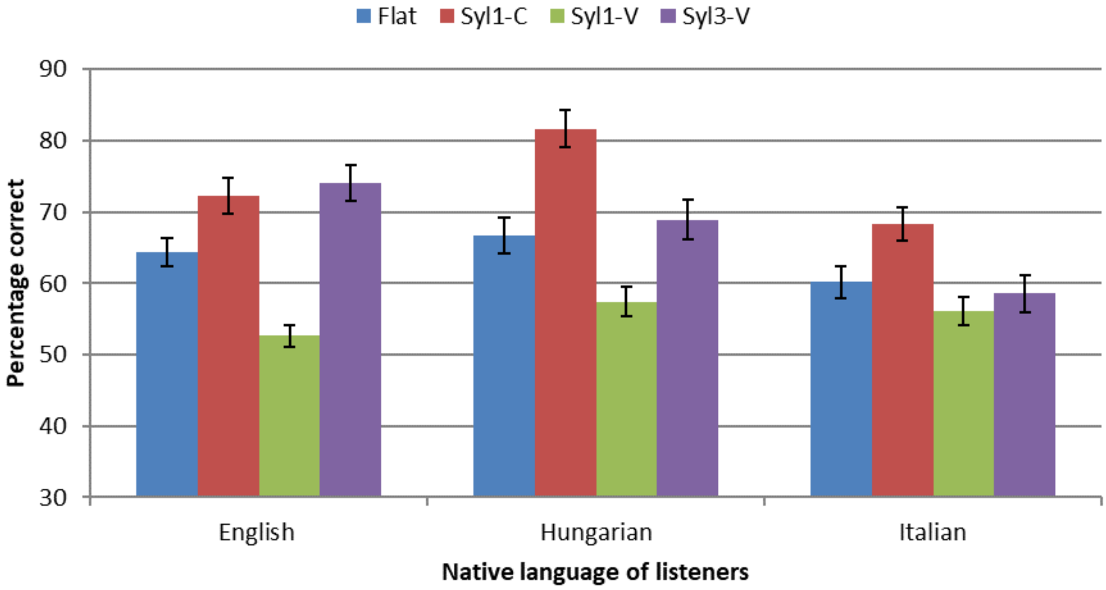


Figure 1. Mean correct recognition of nonwords after six minutes' exposure to an artificial language comprised of four trisyllables: *nudopa*, *subako*, *bitusa*, *ripolu*. See abstract (paragraph 2) for details of the timing conditions: **Flat**, **Syl1-C**, **Syl1-V**, **Syl3-V**.

[1] S. L. Mattys, L. White, and J. F. Melhorn, 'Integration of multiple speech segmentation cues: a hierarchical framework.', *J. Exp. Psychol. Gen.*, vol. 134, no. 4, pp. 477–500, 2005.
[2] A. Fernald and G. McRoberts, 'Prosodic bootstrapping: A critical analysis of the argument and the evidence', *Signal Syntax Bootstrapping Speech Gramm. Early Acquis.*, pp. 365–388, 1996.
[3] J. Fletcher, 'The prosody of speech: Timing and rhythm', *Handb. Phon. Sci. Second Ed.*, pp. 521–602, 2010.
[4] M. D. Tyler and A. Cutler, 'Cross-language differences in cue use for speech segmentation', *J. Acoust. Soc. Am.*, vol. 126, no. 1, pp. 367–376, 2009.
[5] P. Keating, T. Cho, C. Fougeron, and C.-S. Hsu, 'Domain-initial articulatory strengthening in four languages', in *Phonetic Interpretation: Papers in Laboratory Phonology VI*, J. Local, R. Ogden, and R. Temple, Eds. Cambridge: Cambridge University Press, 2004, pp. 143–161.
[6] H. Quené, 'Segment durations and accent as cues to word segmentation in Dutch', *J. Acoust. Soc. Am.*, vol. 94, no. 4, pp. 2027–2035, 1993.
[7] L. White, S. L. Mattys, L. Stefansdottir, and V. Jones, 'Beating the bounds: Localized timing cues to word segmentation', *J. Acoust. Soc. Am.*, vol. 138, no. 2, pp. 1214–1220, 2015.
[8] A. Content, R. K. Kearns, and U. H. Frauenfelder, 'Boundaries versus onsets in syllabic segmentation', *J. Mem. Lang.*, vol. 45, no. 2, pp. 177–199, 2001.
[9] M. Ordin, L. Polyanskaya, I. Laka, and M. Nespor, 'Cross-linguistic differences in the use of durational cues for the segmentation of a novel language', *Mem. Cognit.*, pp. 1–14, 2017.
[10] E. Reinisch, A. Jesse, and J. M. McQueen, 'Speaking rate affects the perception of duration as a suprasegmental lexical-stress cue', *Lang. Speech*, vol. 54, no. 2, pp. 147–165, 2011.

Effects of coda laryngeal features and stimulus language on perceived vowel duration

Chelsea Sanker
Brown University

Vowels are longer before voiced consonants than before voiceless consonants in many languages, including English (Chen 1970) and Hindi (Durvasula & Luo 2014), but not Telugu (Reddy 1988). Hindi and Telugu also differ in effects of breathy voiced stops; vowels are longer before them than before plain voiced stops in Hindi, but not in Telugu. Various explanations for the voicing effect have been proposed (e.g. Chen 1970, Kluender et al. 1988), but no explanation has been substantiated. Less work examines the breathiness effect.

I present two studies on how coda features affect perceived duration of preceding vowels when the codas themselves are removed. For both Hindi and Telugu stimuli, native English listeners perceive vowels produced before voiced stops as longer than vowels from before voiceless stops, suggesting a possible perceptual origin of voice-induced duration differences. Vowels from before Hindi breathy voiced stops are perceived as even longer, but have the opposite pattern for Telugu stimuli, likely reflecting differences in breathiness assimilation.

Study 1: 24 American English speakers heard isolated vowels extracted from VC forms produced by a Hindi speaker, with the consonant and the transition into it removed, and categorized each vowel as 'long' or 'short', with instructions that they were deciding about duration. They were not trained in the contrast. The original codas were voiced, voiceless, and breathy voiced stops. Items were blocked by vowel quality (/a/, /i/, /u/). Vowel duration was manipulated to create stimuli covering a 10-step duration continuum (129 ms to 252 ms).

Study 2: 24 American English speakers heard isolated vowels extracted from VC forms produced by a Telugu speaker and categorized each vowel as 'long' or 'short' in duration. All parameters other than the language used for the stimuli were the same as for Study 1.

Regression models for effects in Study 1 and 2 are given in Tables 1 and 2, respectively; the effects of duration and original coda voicing are illustrated in Figures 1 and 2.

English speakers can perceive and categorize vowel duration; in both studies, the actual duration of the vowel was a major predictor of whether it was identified as 'long' or 'short'.

Coda voicing was a significant predictor of responses. For both Hindi and Telugu stimuli, listeners gave more responses of 'long' when the original coda had been voiced than when it had been voiceless. Vowels exhibit several acoustic influences of following stops, which may be contributing to the differences in perceived duration; within the stimuli for both studies, vowels before voiced stops had lower F1 and less jitter than vowels before voiceless stops.

Coda breathiness was a significant predictor of responses for the Hindi stimuli; listeners gave more 'long' responses when the original coda had been a breathy voiced stop than when it had been plain voiced. In contrast, for the Telugu stimuli, vowels from the breathy voiced environment elicited fewer 'long' responses. The different effects are likely due to differences in realization; vowels before breathy voiced stops are breathy in Hindi, reflected in spectral tilt and harmonics-to-noise ratio (HNR), but not in Telugu.

Vowel quality also matters; high vowels, which are shorter in production than low vowels, elicited more 'long' responses, suggesting listeners' compensation for expected durations.

Vowels produced before voiced codas are perceived as longer than vowels produced before voiceless codas, which suggests a possible perceptual pathway for the development of voicing-conditioned vowel duration, particularly because the perceptual difference is present even in stimuli from a language which lacks the voicing effect. Coda breathiness has different effects based on stimulus language, paralleled by differences in realization; vowels assimilate to breathiness of coda stops only in Hindi. Greater perceived duration of these Hindi vowels suggests that vowel breathiness increases perceived length, providing a possible explanation for why they are often longer than modal vowels (cf. Gordon & Ladefoged 2001).

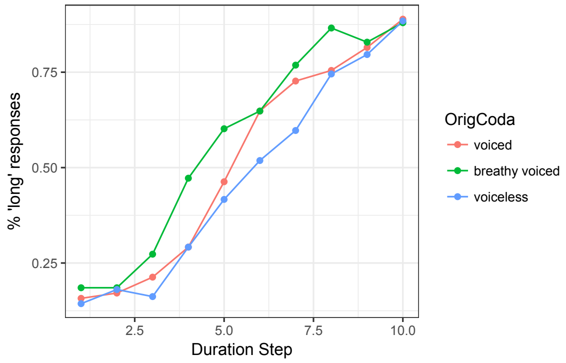


Figure 1. *Study 1 : 'long' responses, by coda features (Hindi stimuli)*

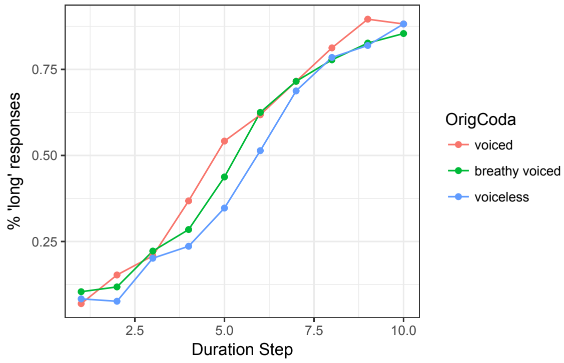


Figure 2. *Study 2 : 'long' responses, by coda features (Telugu stimuli)*

	Estimate	Std. Error	Z value	P value
(Intercept)	-2.58	0.16	-15.9	< 0.0001***
DurationStep	0.48	0.013	38.0	< 0.0001***
OrigCoda-Voiceless	-0.23	0.074	-3.12	0.0018***
OrigCoda-BreathyVoiced	0.35	0.074	4.65	< 0.0001***
Vowel-i	0.22	0.074	2.97	0.003**
Vowel-u	-0.052	0.074	-0.71	0.48
ResponseTime	-0.023	0.057	-0.41	0.68

Figure 1. *Generalized linear mixed model for 'long' responses in Study 1 (Hindi stimuli). 270 trials per participant. Intercept: OrigCoda = Voiced; Vowel = /a/*

	Estimate	Std. Error	Z value	P value
(Intercept)	-3.08	0.18	-117.2	< 0.0001***
DurationStep	0.55	0.017	33.0	< 0.0001***
OrigCoda-Voiceless	-0.40	0.094	-4.2	< 0.0001***
OrigCoda-BreathyVoiced	-0.19	0.094	-2.0	0.044*
Vowel-i	0.27	0.94	2.8	0.0045**
Vowel-u	0.24	0.94	2.6	0.0096**
ResponseTime	0.039	0.041	0.94	0.35

Figure 2. *Generalized linear mixed model for 'long' responses in Study 2 (Telugu stimuli). 180 trials per participant. Intercept: OrigCoda = Voiced; Vowel = /a/*

[1] Chen, M. 1970. Vowel length variation as a function of the voicing of the consonant environment. *Phonetica* 22, 129-159.

[2] Durvasula, K. & Luo, Q. 2014. Voicing, aspiration, and vowel duration in Hindi. In S. Ohlsson & R. Catrambone (eds.), *Proceedings of meetings on acoustics*, 060009. Acoustical Society of America.

[3] Gordon, M. & Ladefoged, P. 2001. Phonation types : A cross-linguistic overview. *Journal of Phonetics* 29, 383-406.

[4] Kluender, K.R., Diehl, R.L., & Wright, B.A. 1988. Vowel-length differences before voiced and voiceless consonants: An auditory explanation. *Journal of Phonetics* 16, 153-169.

[5] Reddy, K.N. 1988. The duration of Telugu speech sounds: An acoustic study. *IETE Journal of Research* 34.1, 57-63.

“Similar or different?” – The phonetics and phonology of similarity in non-native vowel perception

Nadja Kerschhofer-Puhalo[1]
[1]*University of Vienna*

Due to the high dynamics and variability in the speech signal, perceptual vowel identification cannot be predicted directly from the acoustic properties of the signal but is determined by the listener’s attentional tuning to specific acoustic cues and perceptual dimensions varying language-specifically and even individually. Acoustically different vowel sounds are perceived at some cognitive level as more or less “*similar*” or “distinct” from others, belonging either to the same or to different *categories*. In experimental identification tasks, listeners‘ perception of acoustically different sounds as belonging to the ”same“ or a ”different” category are conditioned by signal-inherent as well as external factors from the listeners’ language knowledge, context and expectations. The categorization of sounds is based on relationships of “*similarity*” between single items and mental categories of a given language. Language-specific or more general physical and cognitive *biases* determine the categorization of a given input as “*similar*” to items belonging to mental categories. *Similarity* has been used as one of the central concepts in many models of second language speech perception and acquisition (e.g. Best 1995; Flege 1987, 1995; Kuhl 1992, 1993). Yet, the operationalization of this construct has so far not deserved sufficient attention.

Based on response material from a prior study on non-native vowel identification (Kerschhofer-Puhalo 2014, in print), this presentation will discuss empirically grounded ways of *operationalizing perceptual similarity* in L2 in terms of (1) *acoustic similarity* scores and (2) *psychological similarity* and distances. In a vowel identification task, 15 German vowel types were presented in non-words with differing consonantal context to 173 L2 learners of German from 10 L1 subsamples (+ a native control group). Participants were asked to match the input stimuli with response categories, which consisted of all full vowel categories of German. The L2 listeners’ responses were summarized in confusion matrices and were – together with data from the acoustic analysis of the input stimuli – subject to higher level statistical analysis. Multidimensional Scaling (MDS) was applied to visualize (intra-lingual) similarity of German vowel phonemes in a geometric 2- or 3-dimensional spatial representation of the L2 vowel space (Shepard 1972; Terbeek 1977; Johnson 2012). Several previous experiments using MDS solutions have postulated a high correspondence between spatial distances in MDS solutions and acoustic-phonetic properties such as vowel formants (Kewley-Port & Atal 1989; Iverson & Kuhl 1995; Fox, Flege & Munro 1995; Francis & Nusbaum 2002). Alternatively to more traditional mono-directional conceptions of similarity between L1 and L2 sounds and the claim that statistical correlations with acoustic properties are sufficient to understand *perceived similarity in L2*, we favour a *cross-linguistic influence*-approach focussing on *biases* associated with properties of *stimuli* (acoustic-phonetic) as well as *responses* (phonological) to account for ease and difficulty, preferences and avoidance in L2 perception experiments. A cross-language comparison of the L1 subsamples shows that – rather than predicting perceptual similarity directly from acoustic phonetic properties – perceptual *similarity* s_{ij} between vowel categories of the target language has to be modelled as the result of the complex interaction of (1) *phonetic proximity* p_{ij}, (2) *stimuli* biases b_i and (3) response *biases* b_j ($s_{ij} = p_{ij} * b_i * b_j$). We will show that *biases* vary according to characteristics of the acoustic signal, the set of stimuli and response categories presented in the experimental setting as well as to the L2 learners’ language experience (in L1, L2, Ln), L2 proficiency and their individual conception of the target language vowel system.

[1] Best, C. 1995. A direct realist perspective on cross-language speech perception. In: Strange, W. (ed.). *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Timonium, MD: York Press, pp. 171–204.
[2] Flege, J. 1987. The production of “new” and “similar” phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics*, 15, pp. 47–65.
[3] 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, pp. 233–277.
[4] Fox, R., Flege, J., Munro, M. 1995. The perception of English and Spanish vowels by native English and Spanish listeners: A multidimensional scaling analysis. *JASA*, 97, pp. 2540–2551.
[5] Francis, A. & Nusbaum, H. 2002. Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance*, 28 (2), pp. 349–366.
[6] Hillenbrand, J. 2013. Static and dynamic approaches to vowel perception. In: Morrison, G. & Assmann, P. (eds.). *Vowel inherent Spectral Change. Modern Acoustics and Signal Processing*. Berlin, Heidelberg: Springer, p. 9–30.
[7] Iverson, P. & Kuhl, P. 1995. Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *JASA*, 97, pp. 553–562.
[8] Johnson, K. 2012. *Acoustic and Auditory Phonetics*. 3rd ed. Wiley-Blackwell.
[9] Kerschhofer-Puhalo, N. 2014. *Similarity, Cross-linguistic Influence and Preferences in Non-Native Vowel Perception. An experimental cross-language comparison of German vowel identification by non-native listeners*. Dissertation, University of Vienna, DOI: https://doi.org/10.25365/phaidra.51.
[10] Kerschhofer-Puhalo, N. (in print). Notions of similarity in non-native vowel perception. *Proceedings der P&P14. 14. Tagung Phonetik und Phonologie im deutschsprachigen Raum*, University of Vienna, Sept 6-7, 2018.
[11] Kewley-Port, D. & Atal, B. 1989. Perceptual distances between vowels located in a limited phonetic space. *JASA*, 85, pp. 1726–1740.
[12] Kuhl, P. 1992. Speech prototypes: studies in the nature, function, ontogeny and phylogeny of the ‘centers’ of speech categories. In: Tohkura, Y., Vatikiotis-Bateson, W., Sagisaka, Y. (eds.). *Speech Perception, Production and Linguistic Structure*. Tokyo: Ohmsha, pp. 197–219.
[13] Kuhl, P. 1993. An examination of the ‘perceptual magnet’ effect. *JASA*, 93, p. 2423.
[14] Shepard, R. 1972. Psychological representation of speech sounds. In: David, E. & Denes, P. (ed.). *Human Communication: A Unified View*. New York: McGraw-Hill, pp. 67–113. [http://courses.cit.cornell.edu/ling4476/shepard-on-speech-sounds.pdf]
[15] Shepard, R. 1980. Multidimensional Scaling, Tree-Fitting, and Clustering. *Science*, 210, pp. 390–398. [http://serendip.brynmawr.edu/local/scisoc/emergence/shepard80.pdf]
[16] Terbeek, D. 1977. *A Cross-Language Multi-Dimensional Scaling Study of Vowel Perception*. PhD Dissertation. University of California, Los Angeles. [http://escholarship.org/uc/item/3nx4g138?display=all#page-1]

Keynote speech II

The challenge of word articulation: A neurophonetic view.

Wolfram Ziegler
EKN – Institute of Phonetics and Speech Processing. LMU München, Germany

At a general level of behavioral research, pathology can be revealing about normal function. In speech, more specifically, pathologies of the brain or the peripheral speech organs may inform us about the mechanisms governing normal speech production. Note that the evidence we can hope to establish in clinical speech research depends upon the pathology underlying the medical condition we study.

Neural pathologies afflicting a circumscribed cortical region in the inferior frontal lobe of the left brain, resulting in a syndrome termed *apraxia of speech*, are revealing, because this brain region is considered to store the processes of speech motor planning we have acquired during childhood. As a consequence, investigations into the patterns of apraxic speech impairment may contribute to our knowledge about the architecture of the phonetic planning processes involved in the production of a speaker’s native language. Therefore, apraxia of speech turns out to be particularly relevant for empirical research in phonetics and phonology.

In this talk, I will review some of the research devoted to the speech error patterns observed in individuals with apraxia of speech. A major focus will be on the linguistic factors determining the likelihood that apraxic patients commit errors. I will describe a non-linear probabilistic model based on gestural decompositions of words. The model allows us to predict the relative vulnerability of individual words to apraxic speech impairment by delineating, at the word level, the fracture points of phonetic planning.

Several clinical and non-clinical applications of the model will be outlined to demonstrate its usefulness.

Oral session 4

Pathological speech

Anticipatory VtoV coarticulation in French in several Motor Speech Disorders

Daria D'Alessandro[1], Michaela Pernon[1,2], Cécile Fougeron[1] and Marina Laganaro[3]
[1]*Laboratoire de Phonétique et Phonologie, UMR 7018, CNRS/U. Sorbonne-Nouvelle*
[2]*Dpt des Neurosciences Cliniques, HUG*
[3]*FAPSE, Université de Genève*

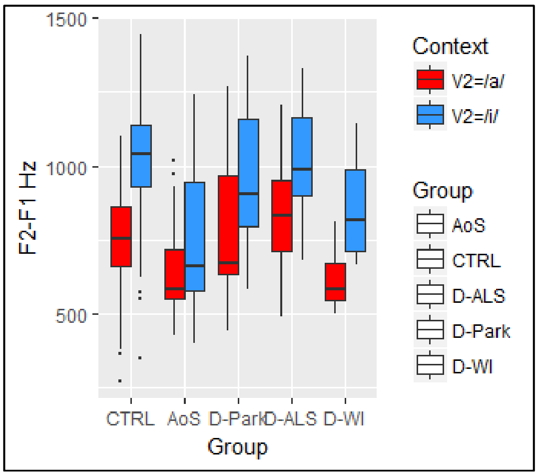
V-to-V anticipatory coarticulation is frequently taken as an indicator of the anticipated planning of an up-coming vowel, and thus the manifestation of a process where context-dependent speech targets or coordinated gestures are planned together before execution (e.g. [1], [2], [3]). In this study, V-to-V anticipatory coarticulation in French is compared in 4 groups of patients presenting different Motor Speech Disorders --Apraxia of Speech (AoS) and 3 types of dysarthrias associated with Amyotrophic Lateral Sclerosis (D-ALS), Parkinson (D-Pk) and Wilson (D-Wl) diseases-- with the aim of getting further insights into the understanding of the control of coarticulation in speech (as in [4], [5], [6]). It is hypothesized that perturbed coarticulatory patterns may arise as a consequence of disruption in the planning and sequencing of speech targets into cohesive speech units, as expected in AoS where speech is often described as 'syllabified' or 'segmentalized'. However, it can also derive from deficits in the motor programming or execution of speech whose consequences are slower movements, restricted displacement and/or perturbed coordination between gestures, as can be encountered in dysarthria. So far, inconsistent results have been found in the literature regarding coarticulatory patterns in different MSD (see [5] & [6]). Moreover, AoS patients are more often compared to patients with aphasia than to patients with dysarthria. More data is thus needed to understand whether deficits linked to distinct cerebral disorders do impact coarticulation, and how they can shed light on the levels and units over which coarticulation is planned.

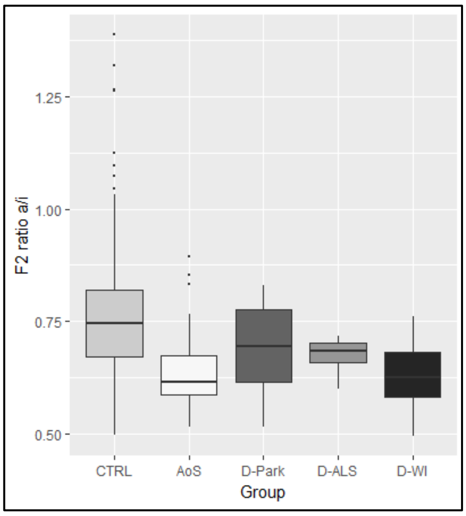
Twenty patients (five per groups) suffering from mild to moderate speech impairments, and 40 healthy control speakers (age and sex matched) were recorded while reading a short text containing twelve disyllabic [pV_1pV_2] French words with V_1=/a/ and V_2=/a/ or /i/. The pivot bilabial consonant /p/ was chosen in order to maximize the effect of V_2 on V_1 shown by a lowering of F1 and a rising of F2, leading to a less compact F2-F1, when /a/ is followed by /i/.

Results show a significant contextual effect of V_2 on V_1 on F2-F1 in all groups of patients with dysarthria and in the control group, but not in the group of patients with AoS (Fig. 1), as found in [8], for instance. Nonetheless, some contextual effect is also present for some speakers in the AoS group, with a certain speaker- and token-dependent variability (Fig.3). In order to compare further the groups, the amount of coarticulation in the /papi/ (V_2= /i/) words was computed as the ratio between the F2 of the first vowel /a/ and the F2 of the following /i/. Reduced amount of coarticulation (more distance between [a] and [i] F2's) is found in both the D-Wl group and AoS groups, compared to the CTRL and other MSD groups (Fig. 2).

Altered coarticulatory patterns (absent or reduced) for the patients in the AoS and D-Wl groups can both be related to deficits in speech coordination and timing, but over different domains. In the Wilson group, reduced coarticulation is associated with longer V_1 and with restricted articulatory movements' displacement, as shown by a reduced acoustic space for all Wilson patients especially in the F1 dimension (linked to reduced jaw displacements). Taken together this could suggest that it is at the segment level that spatio-temporal coordination is altered for these speakers, with a pressure towards articulatory accuracy rather than speed, and potentially less overlap between V_1 and V_2. The AoS group presents longer V_1 targets but also longer intervals between V_1 and V_2, due to either long pivot Cs and/or internal pauses, suggesting an altered phasing between the two syllables in the [pV_1.pV_2] word. Together with the lack of anticipation of the up-coming V_2 during the first syllable, AoS

patients seem to proceed with syllable-sized phonetic plans lacking inter-syllable coordination.
While more speakers per patient groups are needed, and local anticipatory coarticulation remains to be examined, theoretical implications of these results will be discussed at the conference.





Apraxia of Speech (AoS) ; Dysarthia associated with Wilson disease (D-Wl), with Amyotrophic Lateral Sclerosis ((D-ALS); with Parkinson disease (D-Pk) ; healthy control speakers (CTRL)

***Figure 1**. Acoustic compacity of V_1=/a/, measured as F2-F1, per context(V_2=/a/ or V_2= /i/) and per group*

***Figure 2**. Degree of coarticulation, measured by the ratio of F2 in V1/a/ and V2/i/ in /api/ words, per group (smaller ratio = less coarticulation)*

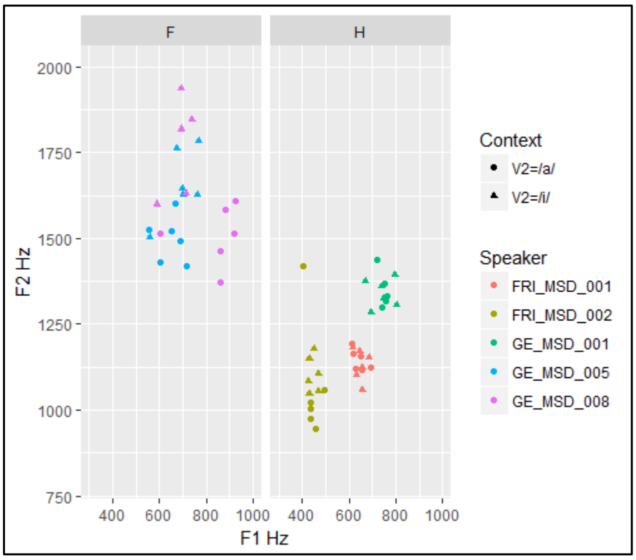


***Figure 3**. F1 and F2 of /a/ according to V2 context (/a/ or /i/) for the 5 patients with Apraxia of Speech (different color per patient, 12 tokens each, F=female, H= male speakers).*

References:

[1] Whalen, D. H. 1990. Coarticulation is largely planned. *Journal of Phonetics*, 18, 3-35.
[2] Bell-Berti, F., Krakow, R. A., Gelfer, C. E., & Boyce, S. E. 1995. Anticipatory and carryover effects: Implications for models of speech production. In F. Bell-Berti & L.J. Raphael (Eds.), Producing speech: contemporary issues for Katherine Safford Harris, 77-98.
[3] Recasens, D. 2002. An EMA study of VCV coarticulatory direction. The Journal of the Acoustical Society of America, 111(6), 2828-2841.
[4] Hertrich, I., & Ackermann, H. 1999. Temporal and spectral aspects of coarticulation in ataxic dysarthria: An acoustic analysis. *Journal of Speech, Language, and Hearing Research*, 42(2), 367-381.
[5] Katz, W.F. 2000. Anticipatory coarticulation and aphasia: implications for phonetic theories. *Journal of Phonetics*, 28, 313-334

[6] Frisch, S. A., Maxfield, N., & Belmont, A. 2016. Anticipatory coarticulation and stability of speech in typically fluent speakers and people who stutter. *Clinical linguistics & phonetics*, 30(3-5), 277-291.
[7] Hardcastle, B., & Tjaden, K. 2008. Coarticulation and Speech Impairment. In. M. J. Ball, M.R. Perkins, N. Müller, S. Howard, eds., The handbook of clinical linguistics. Blackwell.
[8] Ziegler, W., & von Cramon, D. (1986). Disturbed coarticulation in apraxia of speech : acoustic evidence. *Brain and Language*, 29, 34-47.

Cognitive abilities and prosody in question-response interactions: A clinical study

Caterina Petrone[1], Mélody Zira[1], Christelle Zielinski[1,2] and Elisa Sneed German[3]

[1]*Aix-Marseille Université, CNRS, LPL, UMR 7309, Aix-en-Provence, France;*
[2]*Institut of Language, Communication and the Brain, Aix Marseille Université;*
[3]*Université Paul Valéry Montpellier 3, EMMA, Montpellier, France*

In question-response interactions, short latencies (~200 ms) require that listeners grasp relevant information from their interlocutors' questions as soon as it becomes available in order to plan their responses before the ongoing turn is finished [e.g., 1, 2]. However, strategies for speech production planning vary with cognitive abilities. For instance, speakers with a high speed of processing plan their speech in a more incremental way and begin speech more quickly [3].

This study investigates the effects of cognitive abilities in question-response interactions from a clinical perspective, by comparing French healthy controls (HC) and individuals with cognitive disorders related to Multiple Sclerosis (MS). MS is an autoimmune disease characterized by the production of widespread demyelinating lesions in the brain and spinal cord. Cognitive impairment (CI) affects 40–65% patients with MS [4] and includes deficits on specific cognitive capacities involved in speech planning [5].

In a question-response game, we monitored eye movements to lexical competitors (**ca**nard/**ca**non, "duck/cannon") during question comprehension as well as latencies of speech responses. Participants orally replied to 24 trials consisting of a sequence of two prerecorded questions (Q1 and Q2). Q1 asked whether the location of one object was above or below a geometric shape (e.g., *Est-ce que le canard est au dessus du rond?*, "Is the duck above the circle?"), and the following Q2 (e.g., *Et est-ce que le canard/canon est en dessous du carré?* "And is the duck/cannon below the square?") asked the location either of the previously mentioned object (anaphoric condition) or of a new object (non-anaphoric condition). Participants responded to Q2 keeping the form of their response constant (e.g., "No, the duck is above the square"). Questions were all parsed into 3 Accentual Phrases (the basic prosodic unit in French). The object name could either carry a focal accent on the ambiguous syllable or not [6]. We expected for HC that the focal accent on the ambiguous syllable should facilitate a non-anaphoric interpretation, while the lack of the accent should induce higher fixations to the just-mentioned object [7]. Given their cognitive impairment, MS patients were predicted to have more trouble in immediate, parallel processing of intonation information, resulting in smaller effects of prosody on eye movements and increased speech latencies.

A sample of 26 patients of early stages of relapsing-remitting MS (< 12 years) and 18 HC (matched in age, sex, and education) has participated so far in the experiment. (Among the inclusion criteria: no relapses at the time of the study; no history of optic nevritis; optimal vision; absence of dysarthria, assessed through tests from the BECD [8]). Participants underwent a series of standard neuropsychological tests, with the two populations differing only in semantic and phonemic fluency, speed of information processing (WAIS IV), and vocabulary (Mill Hill Vocabulary scale) ($p<0.05$). Logistic models with mixed effects showed a bias for fixating "old" pictures, i.e., that were already referred to in Q1 ($t=-2.3$, $p=0.02$) in both HC and MS patients. For HC, in the non-anaphoric condition, the likelihood of fixations to the already mentioned object was higher when the word was deaccented than when the word was focused ($t=1.9$, $p=0.04$), with phonetic material *preceding* the ambiguous syllable already guiding listeners' interpretation ($t=2.3$, $p=0.01$). No effects of prosody were found in the anaphoric condition. For MS patients, prosody had no effects on eye movements. MS (386 ms) than in HC (202 ms), indicating that cognitive deficits might be strongly disruptive in a real interaction ($t=2.5$, $p=0.01$). Hence, cognitive abilities affect dialogic interactions, with patients at early MS stage having more difficulty both in question comprehension and response planning. This is also in line with the hypothesis that speech production planning is flexible [9]. We aim to further enlarge the number of participants and to investigate the correlations between neuropsychological scores and individual eye movements and verbal performances.

Figures

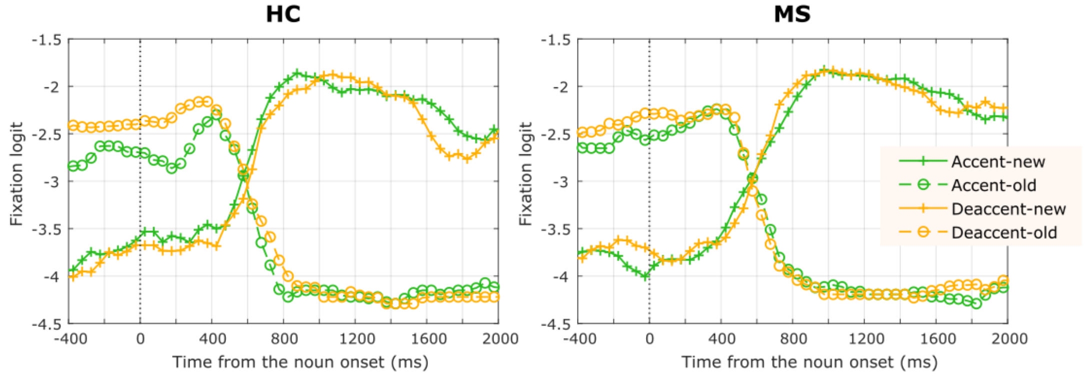


Figure 1. Fixation logits for Q2 in the non-anaphoric condition (canard -> canon): Trials with vs without focal accent are signaled by different colors (green vs. yellow). Fixation logits for Q2 in the non-anaphoric condition (canard -> canon): Trials with vs. without focal accent are signaled by different colors (green vs. yellow). Dots = fixation to the old object, ie, to the object already mentioned in Q1; crosses = fixations to the new object.

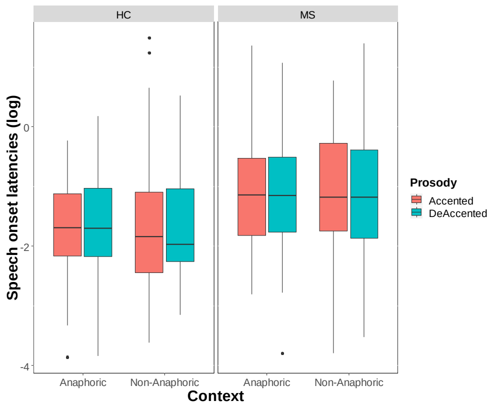


Figure 2. Latency of speech reponses (log) across discourse contexts and prosody in Healhy Controls (left) and MS patients (right).

References

[1] Bögels, S., Magyari, L., & Levinson, S. C. (2015). Neural signatures of response planning occur midway through an incoming question in conversation. Scientific Reports, 5: 12881. doi:10.1038/srep12881.
[2] Barthel, M., Sauppe, S., Levinson, S. C., & Meyer, A. S. (2016). The timing of utterance planning in task-oriented dialogue: Evidence from a novel list-completion paradigm. Frontiers in Psychology, 7. http://dx.doi.org/10.3389/fpsyg.2016.01858
[3] Swets, B., Petrone, C., Fuchs, S. and Krivokapić, J. (2016). Variation in prosodic planning among individuals and across languages. CUNY, 2016, Gainesville (US).
[4] Langdon, D. W. (2011). Cognition in multiple sclerosis. Current Opinion in Neurology, 24(3), 244–249. https://doi.org/10.1097/WCO.0b013e328346a43b
[5] De Looze, C., Moreau, N., Renié, L., Kelly, F., Ghio, A., Rico, A., ... & Petrone, C. (2017). Effects of cognitive impairment on prosodic parameters of speech production planning in multiple sclerosis. *Journal of Neuropsychology*. https://doi.org/10.1111/jnp.12127
[6] Jun S.-A. & Fougeron C. (2000). A Phonological model of French intonation. In A. Botinis (ed.) Intonation: Analysis, Modeling and Technology. Dordrecht : Kluwer AP. 209-242.
[7] 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.
[8] Auzou, P. and Rolland-Monnoury, V. (2006). BECD Batterie d'évaluation clinique de la dysarthrie. Isbergues: Ortho Edition.
[9] Swets, B., Jacovina, M. E., & Gerrig, R. J. (2014). Individual differences in the scope of speechplanning: Evidence from eye movements.Language and Cognition,6, 12–44.

Oral session 5

Variation and modelling

Variation, variability and category overlap in intonation

Georg Lohfink[1], Argyro Katsika[2], Amalia Arvaniti[1]
[1]University of Kent, [2]University of California, Santa Barbara

A major issue in intonation research is modelling the variability of F0 contours while capturing significant generalizations that guide phonological abstraction. Some models ignore variability altogether by dealing with idealized contours [1]; others [2] focus on capturing variability, but at the expense of generalization [3]. The autosegmental-metrical model of intonational phonology (AM, [4]) captures phonological generalizations but has difficulty dealing with variability, as its diagnostics rely on phonetic invariance: tonal targets are the reflexes of underlying tones if they show invariant alignment and scaling. This criterion is at odds with the extent of variability found in natural speech, e.g. [5], and thus requires radical rethinking.

Here we investigate variability in a Greek corpus including 844 tokens of three pitch accents (H*, L+H* and H*+L) all of which appear in utterance-final, nuclear position. The data were collected from Greek speakers (10F, 3M) reading four repetitions of dialogues designed to elicit the three accents on test words with varied stress; see (1). For each test word, the three-syllable interval ending at the offset of the stressed syllable, underlined in (1), was marked and its F0 extracted using STRAIGHT [6]. The F0 curves of this interval underwent Functional Principal Component Analysis (PCA). PCA returns the most dominant modes of variation in functional form, called Functional Principal Components (PCs). Every input curve then receives a coefficient for identified PCs, representing the contribution of each PC to that curve's shape. PCA, and the coefficients for PC1 and PC2 (henceforth *scores*) were statistically analysed together with duration, using linear mixed effects models in R [7, 8] with accent type, stress position, and duration as fixed factors, and speaker as random intercept.

PC1 and PC2 (Fig.1) captured 87.7% of the variance in the corpus (Fig.1), with PC1 reflecting differences primarily in peak height, and PC2 reflecting a combination of contour *shape* and peak alignment (position of the peak in the three-syllable window). Differences in PC1 and PC2 were sufficient to distinguish each accent from the other two (see Fig. 2a), despite the observed overlap between PC scores (Fig. 2b, 2c). Crucially, both PCs were needed to distinguish the accents: PC1 scores were significantly higher for L+H* and H*+L as compared to H*, while PC2 scores were significantly higher only for L+H* as compared to H*. Stress position affected PC1 and PC2 but with H* being less affected than H*+L and L+H*. PC scores also interacted with duration: PC1 decreased with increased duration for both H* and L+H*, while this effect was only observed with respect to L+H* in relation to PC2 (see Fig. 3).

These results showcase the usefulness of data-driven parametrization of F0 curves using PCA, and have consequences for established practices in the study of intonation, especially the invariance criterion. First, they show that variability is widespread but its extent is accent-specific (e.g. H* is less variable than L+H* and H*+L). Second, the results indicate that tonal alignment should not be prioritized over scaling, and that the two are not independent of each other; e.g. PC1 reflects primarily scaling but also differences in peak alignment. Further, accentual contrasts are shown to rely on a number of phonetic dimensions (scaling, peak alignment, segmental duration, curve shape), some of which are more important for some accents than others (e.g. curve shape for L+H*). Finally, some parameters are in trading relations, such as duration and scaling (for PC1), and duration and shape (for L+H* with respect to PC2). Overall, the results suggest that the established research focus on localized F0 targets and invariance as criteria for the phonological status of tonal events risks positing categories that are too fine-grained and capture phonetic variability rather than essential contrasts. Instead the results argue in favour of treating tonal events similarly to segments, i.e. as being expressed by a number of phonetic parameters that show variability and are in trading relationships with each other. A new model based on these principles will be presented and discussed.

(1) **H*** — *What's this?* [laðoˈlemono] *Oil-lemon-sauce.*

H*+L — *What should I do with all these lemons?* [lemoˈnaða] *Lemonade.*

L+H* — *Did you say their son has brown eyes?* [ɣalaˈna] *Blue!*

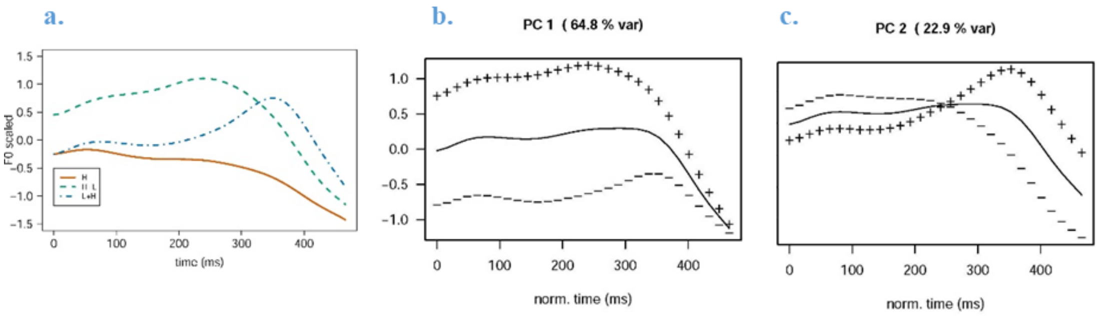


Figure 1. *Average F0 contours of pooled data (a); PC1 (b) and PC2 (c) curves modelling the data in (a) [solid black line = mean curve; + = higher PC scores; - = lower PC scores].*

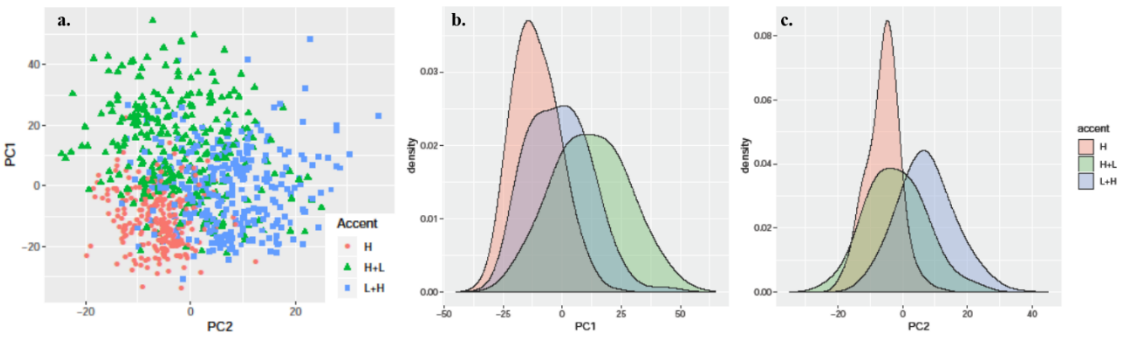


Figure 2. *Scatterplot of PC1 * PC2 by accent (a); density plots for PC1 (b) and PC2 (c) by accent.*

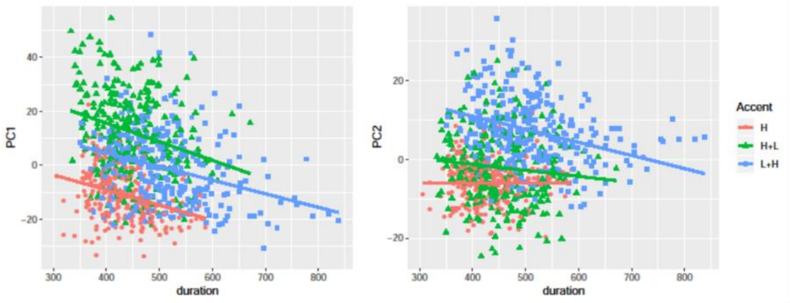


Figure 3. *Scatterplots of duration by PC1 (left) and PC2 (right) separately for each accent type.*

[1] 't Hart, J., Collier, R., Cohen, A. 1990. *A Perceptual Study of Intonation*. Cambridge: CUP.

[2] Xu, Y. 2005. Speech melody as articulatorily implemented communicative functions. *Speech Communication* 46(3-4): 220-251.

[3] Arvaniti, A., Ladd, D. R. 2009. Greek wh-questions and the phonology of intonation. *Phonology* 26: 46-63.

[4] Ladd, D. R. 2008. *Intonational Phonology*. Cambridge: CUP.

[5] Arvaniti, A. 2016. Analytical decisions in intonation research and the role of representations: lessons from Romani. *Laboratory Phonology* 7(1): 6, pp. 1–43.

[6] Kawahara, H., et al. 2005. Nearly defect-free F0 trajectory extraction for expressive speech modifications based on STRAIGHT. *Proceedings of Interspeech 2005*.

[7] Gubian, M., Torreira, F., Boves, L. 2015. Using functional data analysis for investigating multidimensional dynamic phonetic contrasts. *Journal of Phonetics* 49: 16-40.

[8] Douglas B., Maechler, M., Bolker, B., & Walker, S. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1): 1-48.

[9] R Core Team. 2017. *R: A Language and Environment for Statistical Computing*. https://www.R-project.org/.

The contribution of dynamics to the perception of tonal alignment

Leonardo Lancia[1], Cristel Portes[2]
[1]*Laboratoire de Phonétique et Phonologie (CNRS – Sorbonne Nouvelle), Paris, France*
[2]*Aix Marseille Univ, CNRS, LPL, Aix-en-Provence, France*

It is still quite unclear how listeners combine the different features of the f0 signal in the perception of the intonational categories. For example, because the durations of phonologically pertinent f0 movements are weakly correlated to their extension [1], in the perception of pitch accents listeners should be able to factor out the effects of segmental durations on the slopes of f0 rises and falls. Still, the perception of an f0 peak seems to be affected by differences in the slope of the rising (or falling) movement both when these are due to higher starting (or ending) f0 values [2] and when these are due to differences in the curvature of the f0 movement [3]. This sensitivity does not affect the attribution of basic phonological traits (e.g. high/low) to f0 movements, but the perception of their anchoring to the segmental material, i.e, it affects the perception of tonal alignment [2, 3]. To explain these intricate behavioural patterns, it has been proposed [2, 3] that the perceptual location of tonal events depends in some way on the average location of high f0 values in the speech signal. However, this explanation is unable to account for asymmetries in the effects of f0 shapes. Indeed the perception of tonal alignment is more sensitive to changes in the rise preceding a high tone than to changes in the following fall [2].

We present a model that overcomes this limitation by building on the role played by perceptual dynamics in the mapping of continuous f0 values onto abstract intonational events. More specifically, we conceived a simple variant of the Drift Diffusion Model (see [4] for a review) and we conducted a simulation experiment replicating the results obtained in published works ([2] and [3]). In the model, the current state of the perceptual system is characterized by the value of the variable y in the following way: if y exceeds a positive threshold, an upward pitch movement is perceived, while if it crosses a negative threshold, a downward pitch movement is detected. An f0 peak is perceived when both an upward f0 movement and a downward f0 movement have been detected. Governed by the following law of change of y: $\dot{y} = c\dot{f}0 - \lambda y + \xi$, the model states that at a given instant the derivative of y (i.e. its amount of change) is equal to the current rate of change of the f0 signal weighted by the free parameter c (determining the sensitivity of the perceptual system to changes in *f0*), minus the current value of y weighted by the parameter λ (determining how fast the perceptual system returns to a neutral state when *f0* stops changing), plus a random term ξ (representing perceptual noise).

By showing that the proposed model replicates the known effects of f0 shapes on the perception of tonal alignment (see Figure 1) we demonstrate the relevance of dynamics in the explanation of the mechanisms underlying intonation perception. Disregarding this dimension may lead to faulty interpretations of listeners' perceptual behaviour. For example, our model explains the relation between the shape of the f0 curve and the perceived tonal alignment without assuming that listeners explicitly store information on the shape of f0 trajectories. In that, our account resonates with previous models based on the average location of high f0 values, whose partial success can easily be explained as resulting from the accumulative nature of the perceptual drift process. The simplicity of dynamical models like the drift-diffusion model proposed here is due to the fact that they are specific to the constraints of a particular perceptual task and, as shown by a prolific research tradition (see [4] for a review), it guarantees that model parameters are easy to interpret. Additional work is required to account for the perception of more complex f0 patterns and for the effects of spectral and amplitude changes on the perception of f0 changes (e.g.: [5]), and to show how the observed low-dimensional dynamics emerge from the complex interactions between the different factors affecting the perception of intonation.

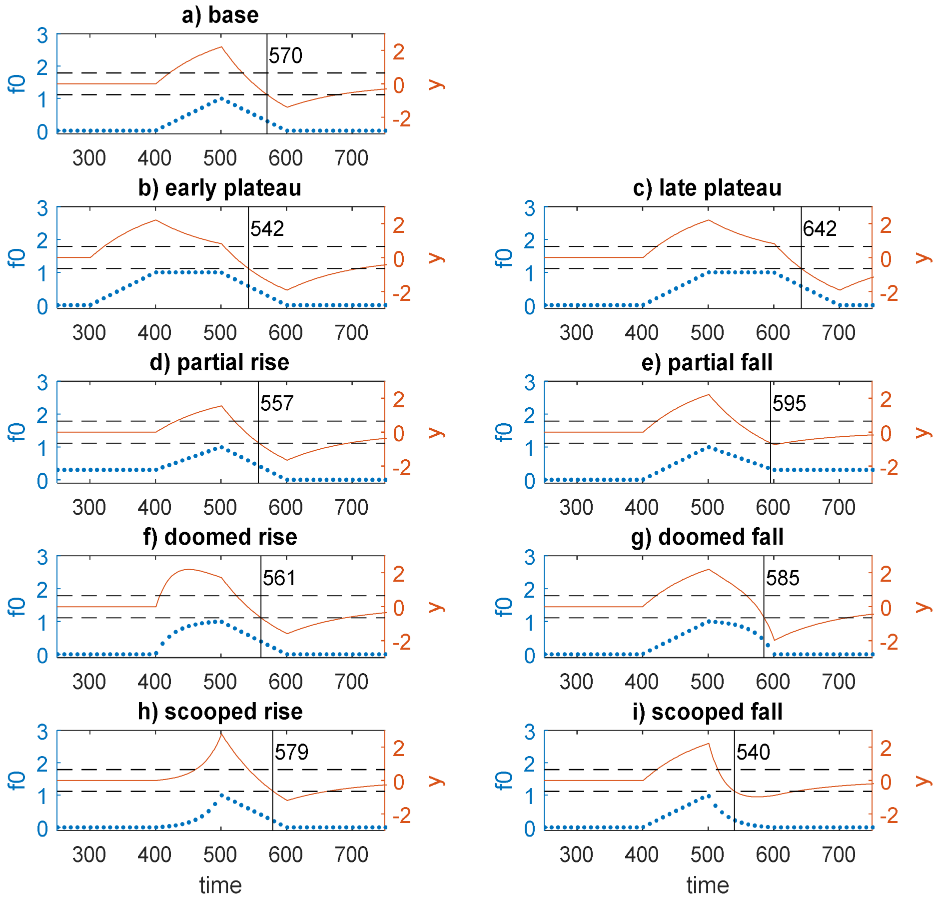


Figure 1. *Behaviour of the model during the presentation of several versions of the basic f0 contour depicted in panel a. The dotted blue curves represent the f0 contour (values in arbitrary units on the left axis) and the continuous red curves represent the state of the perceptual system (values in arbitrary units on the right axis). The horizontal dashed lines represent the thresholds that trigger the detection of pitch movements. The vertical lines indicate the time-point when the fall is detected and the f0 peak perceived. The numbers at their right side indicate the corresponding point in time. Panel a: simulated perception of a symmetric f0 peak. The following modifications to this basic shape anticipate the perception of the f0 peak: early plateau (panel b), partial rise (panel d), doomed rise (panel f), scooped fall (panel i). The perception of the f0 peak is delayed by the following modifications: late plateau (panel c), partial fall (panel e), doomed fall (panel g), scooped rise (panel h).*

[1] Ladd, D. R., Faulkner, D., Faulkner, H., & Schepman, A. (1999). Constant "segmental anchoring" of f0 movements under changes in speech rate. *The Journal of the Acoustical Society of America*, *106*(3), 1543-1554.

[2] D'Imperio, M. (2000). The role of perception in defining tonal targets and their alignment (Doctoral dissertation, The Ohio State University).

[3] Barnes, J., Veilleux, N., Brugos, A., & Shattuck-Hufnagel, S. (2012). Tonal Center of Gravity: A global approach to tonal implementation in a level-based intonational phonology. *Laboratory Phonology, 3*(2), 337-383.

[4] Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological review, 113*(4), 700.

[5] House, D. (1996). Differential perception of tonal contours through the syllable. In *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP'96* 4, 2048-2051.

Prosodic variation in a multi-dimensional dynamic system

Simon Roessig & Doris Mücke
IfL – Phonetik, Universität zu Köln

The research literature in the fields of phonetics and phonology is rife with reports on the variability of prosodic patterns. In particular, one can identify two types of variation [1]: on the one hand categorical variation in the form of a probabilistic mapping of what has been agreed upon as prosodic categories can be found [2]. While, for example, a particular pitch accent type is used to express a pragmatic meaning most of the time, also other pitch accents can be found to be employed for the same function. On the other hand, a great deal of continuous variation within the prosodic categories has been shown. For example, the pitch excursion of an accent can be varied systematically in association with a pragmatic function [3]. It has been suggested that the two types of variation work in symbiosis and can be exploited by speakers flexibly [1].

A growing body of research demonstrates that prosodic variation affects all levels of speech production including the articulatory domain (e.g. [4–6]). Prosodic prominence is expressed by systematic adjustments of the *supra-laryngeal* articulation, i.e. by increasing articulatory effort in prominent syllables to produce longer and larger movements of the tongue, jaw and lips. More recently, it could be shown that systematic variation on the hypo-hyper-articulation continuum is not only modulated by the presence or absence of a pitch accent. Rather, speakers seem to use continuous prosodic variation to directly encode functions like different focus types [7]. Since prosodic variation is so pervasive on all levels of speech, the aim of this study is to bring together the laryngeal-tonal aspects and the supra-laryngeal-articulatory aspects of prosodic variation and propose a modelling account that integrates both. Therefore, we recorded and analysed a corpus of 27 German native speakers with 3-D Electromagnetic Articulography who produced sentences with varying focus structures (broad, narrow, contrastive focus, and background).

Fig. 1 presents results of two parameters from our analysis: (a) shows the results for the tonal onglide – a laryngeal parameter characterising the f0 movement towards the main target of the pitch accent (negative values indicate falling accents, positive values indicate rising accents). The proportion of rising accents increases when going from broad focus to narrow focus, and finally to contrastive focus. The proportion of the pitch accent categories varies systematically between the focus types as well as the medians of the riding distributions indicated by the black dots. The measure of the lip aperture (b) shows a similar picture of continuous variation with increasing values from broad to narrow focus to contrastive focus.

We employ an *attractor-based* modelling approach rooted in the framework of dynamic systems theory. Dynamic systems have proven to be a powerful tool in understanding human cognition in general and patterns of speech in particular [8–11]. While these systems work in a completely continuous environment, they offer the concept of attractors to describe stability and (near-)categorical behaviour. We model the prosodic system as a multi-dimensional attractor landscape, like the one exemplified in Fig. 2, in which the attractors come into being by the patterns of stability and variability on the various dimensions. Crucially however, the system is governed by a *single* control parameter. By scaling this control parameter, the attractor landscape is shaped, attractors become more or less stable or slightly shift their location. Compare the *left* panel of Fig. 2, where the control parameter is 0 to the *right* panel where it is increased to 4: both the probability of rising accents, as well as the magnitude of the tonal onglide and the lip aperture is increased.

Our model integrates tonal and articulatory variation of prosodic prominence. By capturing both, categorical variation as well as continuous variation, it contributes to the understanding of phonetics and phonology as one system [12].

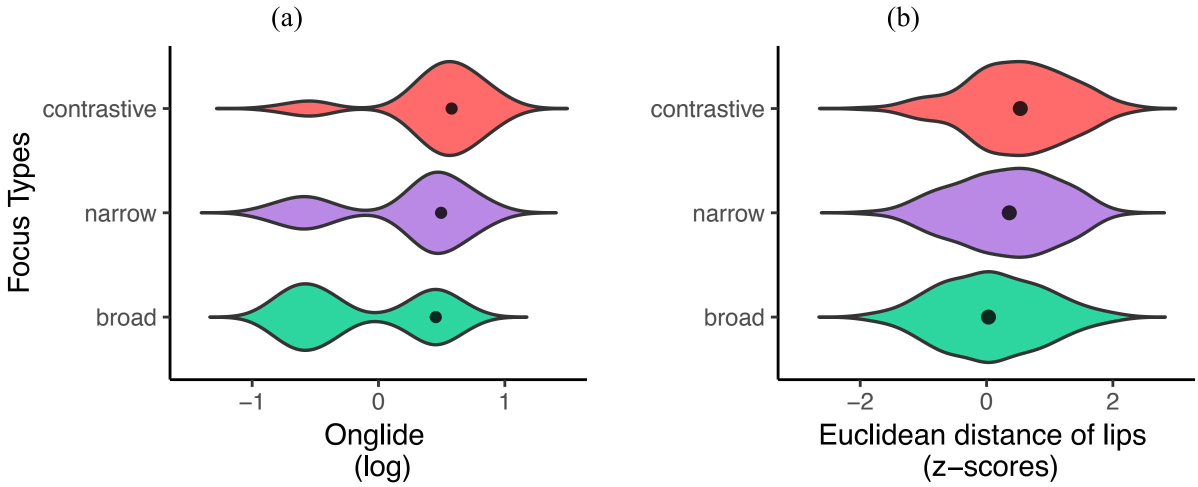


Figure 1. *Results of the measures (a) tonal onglide and (b) lip aperture. Black dots indicate medians of the distribution – in the case of tonal onglide only the rising distributions.*

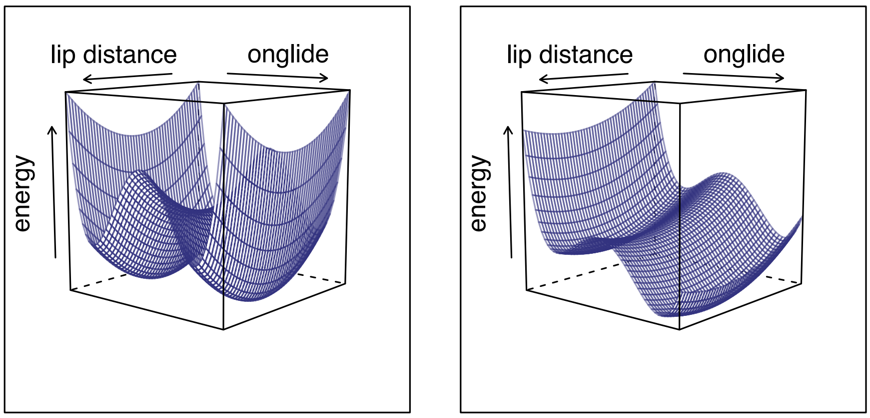


Figure 2. *An example attractor landscapes of the multi-dimensional modelling approach with varying control parameter values k: left k = 0; right k = 4.*

1. Ladd DR. Simultaneous structure in phonology. Oxford: Oxford University Press; 2014.
2. Roettger TB, Cole J. Mapping prosody onto meaning – the case of information structure in American English. 2018. Preprint: https://psyarxiv.com/tevd6/
3. Grice M, Ritter S, Niemann H, Roettger TB. Integrating the discreteness and continuity of intonational categories. J Phon [Internet]. 2017;64:90–107. Available from: http://www.sciencedirect.com/science/article/pii/S0095447017300566
4. Harrington J, Fletcher J, Beckman ME. Manner and place conflicts in the articulation of accent in Australian English. In: Broe M, editor. Papers in laboratory phonology V: Acquisition and the Lexicon. Cambridge: Cambridge University Press; 2000. p. 40–51.
5. Cho T, M. McQueen J. Prosodic influences on consonant production in Dutch: Effects of prosodic boundaries, phrasal accent and lexical stress. J Phon. 2005;33(2):121–57.
6. Mücke D. Dynamische Modellierung von Artikulation und prosodischer Struktur: Eine Einführung in die Artikulatorische Phonologie. Berlin: Language Science Press; 2018.
7. Mücke D, Grice M. The effect of focus marking on supralaryngeal articulation - Is it mediated by accentuation? J Phon. 2014;44:47–61.
8. Spivey M. The Continuity of Mind. New York: Oxford University Press; 2007.
9. Gafos AI, Benus S. Dynamics of Phonological Cognition. Cogn Sci [Internet]. 2006;30(5):905–43. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1207/s15516709cog0000_80
10. Tuller B, Case P, Ding M, Kelso JAS. The nonlinear dynamics of speech categorization. Vol. 20, Journal of Experimental Psychology: Human Perception and Performance. 1994. p. 3–16.
11. Browman CP, Goldstein L. Articulatory phonology: an overview. Phonol Yearb. 1986;3:219–52.
12. Port RF. Dynamical Systems Hypothesis in Cognitive Science. In: Encyclopedia of Cognitive Science. Nature Publishing Group; 2002. p. 1027–32.

Poster session 3

Prenuclear L*+H leads to the activation of alternatives in German

Bettina Braun[1] and Mará Biezma[1,2]
[1]Department of Linguistics, University of Konstanz, Germany, [2]U Mass Amherst, MA

Nuclear accents have received a lot of attention in terms of the prosody-semantics interface, such as information structure/status [1-3], prominence [4, 5] and processing [6, 7]. Prenuclear accents have received less attention in the literature. Semantically, they are described as ornamental [8], serving mostly a rhythmic purpose [9] and listeners don't seem to process them as efficiently as nuclear accents [10, 11]. We tested whether this also holds for prenuclear L*+H in German, an accent which is used to signal contrastive topics CTs [12, 13], among others.

Regarding speech processing, certain nuclear accent types have the potential to make unmentioned, contrastively related alternatives salient to the listener [14, 15]. For instance, Braun, et al. [15] tested German listeners who heard declarative utterances (e.g., *The swimmer wanted to put on flappers*) and watched displays that depicted four printed words: one that was contrastively related to the subject noun (e.g., *diver*), one that was non-contrastively related (e.g., *pool*), the object (e.g., *flappers*), and an unrelated distractor. One experiment compared a nuclear L+H* accent on the subject (indicating a subject focus) to a prenuclear L+H* on the subject (indicating a broad focus). There were more fixations to the contrastive associate when the subject was produced with a nuclear L+H* accent compared to a prenuclear L+H*, suggesting that nuclear L+H* accents activate alternatives [cf. 1]. A nuclear H+L* did not lead to the activation of alternatives, supporting its status as an accent that marks givenness [3].

There is no study on whether prenuclear accents can activate contrastive alternatives. We conducted two visual-word eye-tracking experiments, similar to [15]. In Experiment 1, we compared prenuclear L*+H accents (idealized accent for CTs, with L in tonic and H in post-tonic syllable followed by a high plateau) to prenuclear L+H* accents (idealized accent for broad focus, with L in pre-tonic and H in tonic syllable followed by an f0-dip), following the same procedure for testing and analysis. In Experiment 2 we manipulated the f0-contours of the stimuli in both conditions to match the f0-minima, f0-maxima and f0-excursions of the prenuclear rises to test whether listeners are influenced more by pitch accent type (signaled by the alignment contrast: L*+H vs. L+H*) or by the emphasis signaled by the f0-excursion.

We tested 40 native speakers of German in both Experiments. The results showed that participants fixated the contrastive associate more in the prenuclear L*+H than in the prenuclear L+H* condition in both experiments. In Experiment 1, the difference was significant in the time window 500-600ms after the onset of the subject-NP ($p = 0.005$); the difference approached significance in the time window 600-700ms (see Table 1). Given the time it takes to launch a saccade (150-250ms), cf. Matin, et al. [17], the effect is driven by acoustic information of the subject noun (which starts 200ms and ends 600ms after the onset of the utterance). In Experiment 2, the effect was significant in a later time window, 700-800ms after utteance onset ($p = 0.03$), see Table 1 but the interaction between experiment and intonation condition was not significant in any of the analysis windows (all p-values > 0.2)

Our data show that prenuclear L*+H accents lead to a stronger temporary activation of contrastive alternatives than prenuclear L+H*. This difference persists even if L*+H and L+H* have the same f0-excursion, although the effects occur slightly later in this case. We therefore argue that it is the specific pitch accent type (prenuclear L*+H) that leads to the activation of alternatives and not its phonetic implementation (larger f0-excursion often associated with this contrast in accent types [13]). The results have important implications for theories of contrastive topics (whether they are formalized as an independent information-structural category or hierarchical (as focus nested within topic), cf. [18], [19], [20]) and the relation between phonetics, phonology and processing.

	100 - 200	200-300	300-400	400-500	500-600	600-700	700-800
Exp 1	p = 0.1	p = 0.1	p = 0.07	p = 0.1	**p < 0.005**	p < 0.07	p = 0.3
Exp 2	p = 0.05	p = 0.5	p = 0.7	p = 0.9	p = 0.2	p = 0.1	**p = 0.03**

Table 1. Summary of p-values for the effect of intonation contour on the fixation to contrastive alternatives in subsequent 100ms time windows. The onset and offset of the subject noun are at 200 and 600ms after utterance onset, respectively.

References

[1] Kügler, F. & Gollrad, A. 2015. Production and perception of contrast: The case of the rise-fall contour in German. *Frontiers in Psychology* 6.

[2] Röhr, C. T. & Baumann, S. 2011. Decoding information status by type and position of accent in German. Proceedins of the *17th International Congress of Phonetic Sciences*, (Hong Kong), 1706-1709.

[3] Baumann, S., 2005. The Intonation of Givenness - Evidence from German, PhD Thesis, Philosophische Fakultäten, Universität des Saarlandes, Saarbrücken.

[4] Ayers, G., 1996. Nuclear Accent Types and Prominence: Some Psycholinguistic Experiments, PhD Thesis, Department of Linguistics, Ohio State University, Columbus.

[5] Baumann, S. & Röhr, C. T. 2015. The perceptual prominence of pitch accent types in German. Proceedins of the *18th International Congress of the Phonetic Sciences*, (Glasgow, UK).

[6] Husband, M. E. & Ferreira, F. 2016 The role of selection in the comprehension of focus alternatives. *Language, Cognition and Neuroscience* 31(2), 217-235.

[7] Weber, A., Braun, B., & Crocker, M. W. 2006. Finding referents in time: Eye-tracking evidence for the role of contrastive accents. *Language and Speech* 49(3), 367-392.

[8] Büring, D., Intonation, Semantics and Information Structure, in *The Oxford Handbook of Linguistic Interfaces*, G. Ramchand and C. Reiss, Eds. Oxford: Oxford University Press, 2007, pp. 445-474.

[9] Calhoun, S. 2010. The centrality of metrical structure in signaling information structure: A probabilistic perspective. *Language* 86, 1–42.

[10] Kapatsinski, V., Olejarczuk, P., & Redford, M. A. 2017. Perceptual learning of intonation contour categories in adults and 9- to 11-year-old Children: Adults are more narrow-minded. *Cognitive Science* 41(2), 383-415.

[11] Röttger, T. B. & Cole, J. 2018. On the (ir)relevance of prenuclear accents for intonational learning. Proceedins of the *16th Conference on Laboratory Phonology* (Lisbon, Portugal).

[12] Büring, D. 1997. *The Meaning of Topic and Focus: The 59th Street Bridge Accent*. London: Routledge.

[13] Braun, B. 2006. Phonetics and phonology of thematic contrast in German. *Language and Speech* 49(4), 451-493.

[14] Braun, B. & Tagliapietra, L. 2010. The role of contrastive intonation contours in the retrieval of contextual alternatives. *Language and Cognitive Processes* 25, 1024-1043.

[15] Braun, B., Asano, Y., & Dehé, N. 2018. When (not) to look for contrastive alternatives: the role of pitch accent type and additive particles. *Language and Speech* advance online publication.

[16] McQueen, J. M. & Viebahn, M. 2007. Tracking recognition of spoken words by tracking looks to printed words. *The Quarterly Journal of Experimental Psychology* 60(5), 661-671.

[17] Matin, E., Shao, K. C., & Boff, K. R. 1993. Saccadic overhead: Information-processing time with and without saccadic overhead. *Perception & Psychophysics* 53(4), 372-380.

[18] Wagner, M. 2012. Contrastive topics decomposed. *Semantics and Pragmatics* 5, 1-54.

[19] Steedman, M. 2000. Information structure and the syntax-phonology interface. *Linguistic Inquiry* 31(4), 649-689.

[20] Vallduví, E. & Vilkuna, M., On rheme and kontrast, in *The limits of syntax*, P. Culicover and L. McNally, Eds. New York: Academic Press, 1998, pp. 79-106.

The perception of prosodic cues in Brazilian Portuguese statements and echo-questions: analysis by resynthesis

Luma da Silva Miranda[1], João Antônio de Moraes[1], Albert Rilliard[1,2]
[1]*Federal University of Rio de Janeiro, Brazil*
[2]*LIMSI-CNRS, Université Paris Saclay, France*

This paper presents the results of a study on Brazilian Portuguese statements and echo-questions that measured the relative role of prosodic parameters (melodic movement *versus* intensity and duration) in the perceptual recognition of these intonational contours. The same sentence (*Como você sabe*) was uttered with prosodic variations conveying different speech acts—assertion (meaning "*As you know"*) and echo-question (meaning "*How do you know?"*)—by ten speakers (five males and five females) from Rio de Janeiro (see Fig. 1). Previous studies have shown that the phonological contrast between assertions and echo-questions is performed on the last stressed syllable of the utterance [1, 2]. Using Praat [3], the F0 contours of these two speech acts were manipulated so as to evaluate the role of (i) the direction of F0 movement (rising or falling), and (ii) which syllable carries information on the functional value of these contours. First, a "neutral" stimulus was produced bearing F0 values set at the mean between assertion and echo-question for each syllable. Then, one syllable at a time of this neutral stimulus was modified so as to reproduce the original F0 movement from the statement or the echo-question, for one of the five non-final syllables of the utterance (co/mo/vo/cê/sa). Such modifications of the neutral contour were made either in isolation (each one of the five syllables modified using the two speech act targets — assertion and echo question) or modifying a set of syllables at a time (the pre-stressed and the stressed syllables in the nucleus, the stressed plus post-stressed syllables in the prenucleus and all syllables together). This process resulted in nineteen types of modifications of melodic contours. The role of duration and intensity was tested by transplanting these nineteen F0 stylized contours onto the assertion and echo-question original sentences that serve as phonemic bases, keeping their intensity and duration patterns [4]. There were, thus, thirty- eight synthetic stimuli (nineteen modifications of contours on two base sentences) used for the perceptual experiment. The recognition test was set up using the TP Worken platform [5]. Twenty-four Brazilian listeners had to decide whether each stimulus expresses a statement or a question.

The variation in the proportion of "question" interpretation according to two factors — the *original sentence* (two levels: statement x question) and the type of *stylization* (nineteen levels) —, and their interaction, was analyzed through a logistic regression. The analysis revealed that the interaction between the two factors is not significant. The *original sentence*, which carries differences in duration and intensity, does have a direct influence on the listener's interpretation (an overall 10% increase in "question" interpretation) (see Fig. 2), but that is independent of F0 manipulations, which explains the greatest part of the observed deviance. Moreover, in order to evaluate the perception of each *stylization* as a statement or a question, the "neutral" stimulus was taken as a reference; then the eighteen stylizations were grouped into three sets of stylizations (see Fig. 2): (i) those that are not significantly different from the neutral stylization; (ii) the stylizations with rising '*sa*', falling '*cê*' + rising '*sa*', falling '*você*' + rising '*sa*', falling '*como você*' + rising '*sa*', rising '*cê*' + rising '*sa*', rising '*você*' + rising '*sa*', increased significantly the proportion of "questions" answers, and (iii) the stylizations with falling '*sa*' and rising '*cê*' decreased the proportion of "questions" answers to a significant extent. We conclude that a relevant prosodic cue in Brazilian Portuguese for the distinction between assertion and echo-question is the type of F0 movement in the nuclear region, and especially on the last stressed syllable. Besides, the intensity and duration patterns contributed for the recognition of the functional value of the contours.

(1) Como você sabe

As you know. (assertion) / How do you know? (echo-question)

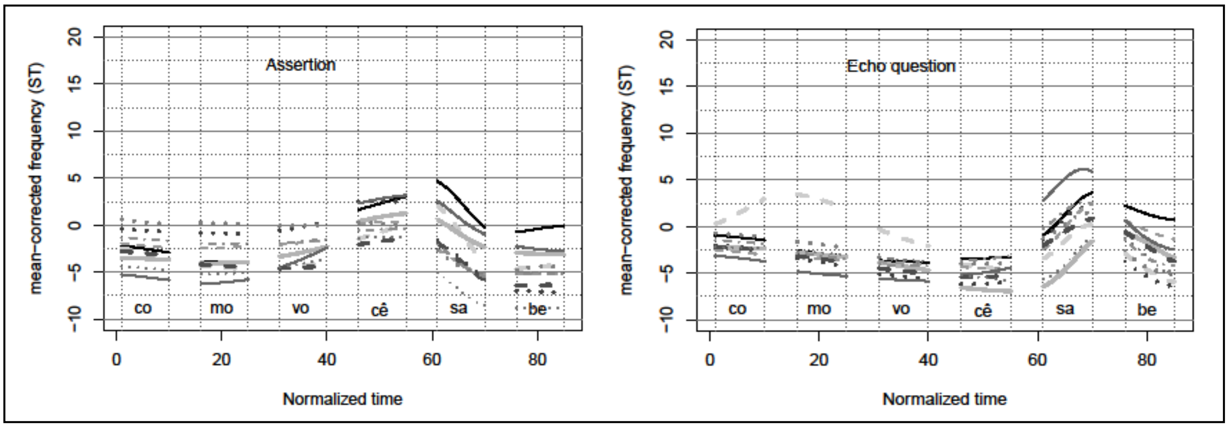


Figure 1. *Plot of the mean F0 values (in ST) of normalized time over each modality: assertion (left) and echo-question (right).*

(2)

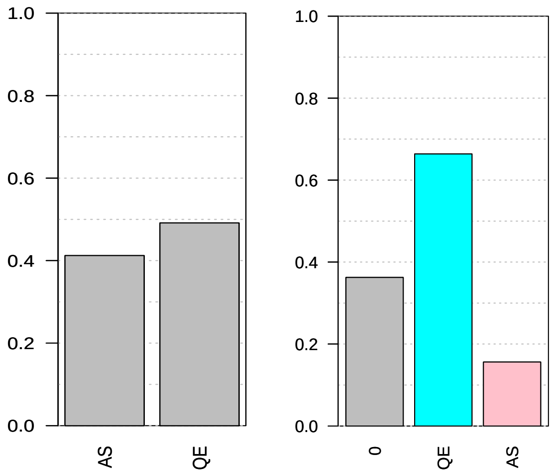


Figure 2. *Mean ratio of interpretations "question" obtained by stimuli of (i) AS (statement) or QE (echo-question) segmental contour (left) and by the three groups of stylizations (right): group 0 (grey), group QE* 'question' *(blue) and group AS* 'assertion' *(pink).*

[1] Moraes, J. 2008. The pitch accents in Brazilian Portuguese: analysis by synthesis. In: SPEECH PROSODY 2008. *Proceedings of the Fourth Conference on Speech Prosody*. Campinas: UNICAMP, pp. 389-397.

[2] Frota, S., Cruz, M., Svartman, F. R. F., Collischonn, G., Fonseca, A., Serra, C. R., Oliveira, P., Vigario, M. 2015. Intonational variation in Portuguese: European and Brazilian varieties. In: Frota, S., & Prieto, P. (Org.). *Intonation in Romance*. 1ed. Oxford: Oxford University Press, v. 1, pp. 235-283.

[3] Boersma, P., & Weenink, D. 2016. *Praat: doing phonetics by computer*. Computer program.

[4] Moraes, J., & Rilliard, A. 2018. Describing the intonation of speech acts in Brazilian Portuguese: methodological aspects. *In*: Feldhausen, I., Fliessbach, J., Vanrell, M. M. (eds.). *Methods in prosody: A Romance language perspective* (Studies in Laboratory Phonology 4). Berlin: Language Science Press, pp. 229-262.

[5] Rauber, A., Rato, A., Kluge, D. C., Santos, G. R., Figueiredo, M. 2012. *TP – Teste de Percepção*. Versão. 3.1.

Variable vowel raising: acoustic analysis of word-initial pretonic mid-vowel /e/ in Brazilian Portuguese

Márcia Cristina do Carmo[1]
[1] *Universidade Estadual de Ponta Grossa (UEPG)*

This work investigates the variable phonetic-phonological process named *vowel raising* of pretonic mid-vowel /e/ in the variety of Brazilian Portuguese (BP) spoken in the countryside of São Paulo State (SP). Our research focus specifically on the word-initial context, e.g. *[i]n.ca.na.dor* ~ *[e]n.ca.na.dor* ('plumber') and *[i]s.pi.nho* ~ *[e]s.pi.nho* ('thorn'), and, therefore, advances in relation to previous studies concerning vowel raising in the same variety (Silveira, 2008;[1] Carmo, 2009;[2] 2013;[3] 2014;[4] Carmo & Tenani, 2013).[5]

Bisol (1981)[6] argues that the linguistic factors that motivate vowel raising of *word-initial* pretonic mid-vowels are different from the ones that condition the application of the process in *word-medial* context, e.g. *m[i].ni.na* ('girl'), and therefore these vowels must be analysed separately. Albeit many studies highlight the protective status of initial syllables concerning alternations (cf. Becker, Nevins & Levine, 2012),[7] studies that describe word-initial pretonic mid-vowels in other varieties of BP (Battisti, 1993;[8] Brandão, Rocha & Santos, 2012)[9] have observed substantial rates of vowel raising regarding word-initial pretonic mid-vowel /e/, e.g. *[i]s.tá.gio* ('internship'). With respect to word-initial pretonic /o/, e.g. *[o].pi.ni.ão* ('opinion'), the application of vowel raising is blocked.

For this research, two experiments were conducted with four participants born and/or raised in the region of São José do Rio Preto city (SP). In the first experiment, the participants read short narratives that elicited contexts of vowel sandhi, with final postonic high vowel [i] and low vowel [a] preceding the pretonic mid-vowel /e/.(1) In the second experiment, the speakers were submitted to repetition of words, preceded by a pause.

In addition to 40 distractors, 20 stimuli were used, previously balanced by a non-parametric (Kruskal-Wallis) test, according to *syllabic structure, frequency, diversity of occurrence* (Tang, 2012),[10] *number of syllables* and *number of letters in the word*. The conditioners considered were *syllabic structure* and *precedent context*. The realization of 20 stimuli in three different contexts (two in the first and one in the second experiment) by four participants totalised 240 items. Due to vowel deletion, one occurrence of */e/s.pe.lho* ('mirror') was discarded from the investigation. The statistical analysis was conducted with the utilisation of PRAAT software (Boersma & Weenink, 2014)[11] and Minitab 17 and Action Stat 3.4 programs.

In a preliminary analysis, there were 69 occurrences (28.9%) of vowel raising of word-initial pretonic mid-vowel /e/. All the 69 occurrences were distributed between two syllabic structures: (i) pretonic /e/ followed by nasal consonant in coda, e.g. *[i]n.fer.mei.ra* ('nurse') (32 occurrences or 53.3% of vowel raising in this syllabic context); and (ii) pretonic /e/ followed by sibilant consonant, as in *[i]s.mal.te* ('nail polish') (37 occurrences or 62.7% of vowel raising in this context). The chi-square tests for association analysis demonstrated (i) a moderate association between the preceding context and the vowel raising; and (ii) a strong association between syllabic structure and the application of the process. In the acoustic analysis, in respect of F1 and F2 values, it was found statistical evidence that the means differ in the different syllabic structures ($p<0.05$) and do not differ in relation to precedent contexts ($p>0.05$). In general, the results highlight the substantial influence of the *syllabic structure* to the vowel raising of word-initial pretonic front mid-vowel.

(1) E.g.: *Pedi que coloque espelho no quarto. Meu pai já foi para a loja, sempre busca espelho lá.* (‘I asked Ø to put a mirror in the bedroom. My father already went to the shop, he always gets mirrors there’).

[1] SILVEIRA, A. A. M. *As vogais pretônicas na fala culta do noroeste paulista*. 2008. 143 f. Dissertação (Mestrado em Estudos Linguísticos) – Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista, São José do Rio Preto, 2008.

[2] CARMO, M. C. *As vogais médias pretônicas dos verbos na fala culta do interior paulista*. 2009. 119 f. Dissertação (Mestrado em Estudos Linguísticos) – Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista, São José do Rio Preto, 2009.

[3] CARMO, M. C. *As vogais médias pretônicas na variedade do interior paulista*. 2013. 249 f. Tese (Doutorado em Estudos Linguísticos) – Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista, São José do Rio Preto, 2013.

[4] CARMO, M. C. As vogais médias pretônicas no noroeste paulista: comparação com outras variedades do Português Brasileiro. *Estudos Linguísticos*. v. 43, n. 1, p. 33-47, 2014.

[5] CARMO, M. C.; TENANI, L. E. As vogais médias pretônicas na variedade do noroeste paulista: uma análise sociolinguística. *Alfa*. v. 57, n. 2, p. 607-637, 2013.

[6] BISOL, L. *Harmonia vocálica: uma regra variável*. 1981. 280 f. Tese (Doutorado em Linguística) – Faculdade de Letras, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1981.

[7] BECKER, M.; NEVINS, A. & LEVINE, J. Asymmetries in Generalizing Alternations to and from Initial Syllables. *Language*. v. 88, n. 2, p. 231-268, jun. 2012.

[8] BATTISTI, E. *Elevação das vogais médias pretônicas em sílaba inicial de vocábulo na fala gaúcha*. 1993. 126 f. Dissertação (Mestrado em Letras) – Instituto de Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 1993.

[9] BRANDÃO, S. F.; ROCHA, F. M. V. & SANTOS, E. R. Vogais médias pretônicas em início de vocábulo na fala do Rio de Janeiro. *Letras & Letras*, Uberlândia, v. 28, n. 1, p. 273-288, jan.-jun. 2012.

[10] TANG, K. A 61 Million Word Corpus of Brazilian Portuguese Film Subtitles as a Resource for Linguistic Research. *UCL Working Papers in Linguistics* 24, 2012.

[11] BOERSMA, P. & WEENINK, D. *Praat:* doing phonetics by computer [Computer program]. Version 5.3.68, retrieved 1 April 2014 from http://www.praat.org/, 2014.

How much in advance can listeners perceive upcoming speech targets? Insights from children and adults

Stella Krüger[1] and Aude Noiray[1,2]

[1]*University of Potsdam, Germany,* [2]*Haskins Laboratories, New Haven, USA*

Coarticulation characterizes the temporal overlap of articulatory gestures for neighboring segments, hence contributing to make speech fast and continuous [1]. In adults, many studies have shown that coarticulation isn´t (solely) a by-product of production but is used by listeners to anticipate upcoming speech material ahead of their acoustic onset [2, 3, 4]. In children, the question remains whether coarticulatory information is used in the same way and to the same extent as in adults. While the literature is replete with studies showing infants´ sensitivity to transitional information (e.g., [5, 6]), the maturational trajectory of speech processing as children transition into more phonologically-driven processing is not well understood.

Previous research investigating vocalic anticipation in speech production revealed that children exhibit greater vocalic prominence over time which results in an earlier initiation of the vowel within syllable and greater temporal overlap of their lingual gestures as compared to adults [7, 8, 9]. Interestingly, this translates into a facilitative vowel identification effect when adults are exposed to child speech (3, 4, 5 and 7 years) as compared to adult´s speech [3]. To the authors´ knowledge, only a few studies have examined this phenomenon further [10, 11] while no study so far investigated vowel anticipation in children.

The present study expands on previous research by examining adults´ as well as children´s anticipatory perception using not only speech samples from adults but also from children of different age groups. The two main hypotheses we tested are that 1) both adults and children would identify upcoming vowel targets earlier in child´s speech as compared to adult´s speech; 2) this effect should be sensitive to gestural compatibility between consecutive segments, resulting into different degrees of perceptual salience of coarticulatory information. More specifically, listeners should anticipate vowels earlier when gestures of adjacent segments recruit different organs and can overlap in time without impeding intelligibility (e.g., /bu/) as compared to cases where the same organ is recruited for both segments (e.g., /du/). However, we expected consonantal effects to be reduced with child stimuli who globally show greater vocalic coarticulation than adults.

To test these hypotheses, we used stimuli previously recorded in a large cross-sectional investigation of coarticulatory patterns in children and adults [7]. We first investigated the ability of 94 adults to detect a target vowel (V_3 = /i:, y:, u:, a:/) ahead of its acoustic onset in short utterances of the form article + nonword ($V_1C_1V_2$ + $\mathbf{C_2V_3}C_3V_4$ with C_2= /b, d/; C_3 = /g/; $V_{2\&4}$= schwa, noted @ from here). A multiple forced choice gating paradigm [12] was designed using the speech of 3-, 5-, and 7-year-olds and adults with 5 temporal gates corresponding to the temporal windows (G1= @ midpoint; G2= @ offset; G3= $\mathbf{C_2}$ midpoint; G4= $\mathbf{C_2}$ offset; G5 = V_3 midpoint). Stimuli were presented in OpenSesame [13] using headphones.

Preliminary results show that accuracy increases as the portion of the signal gets longer (i.e. later gate). Furthermore, performances differ as a function of C_2V_3 pair identity suggesting that perceptual patterns are not uniform across the whole set of vowels for a given consonant. Instead, patterns seem to vary as a function of the gestural specifics for both segments that are coproduced with one another, as found in previous investigations of the coarticulatory patterns themselves [14]. Surprisingly, although listeners showed sensitivity towards coarticulatory differences between CV pairs, a clear age effect was absent. To address these observations in more detail, in-depth analyses are currently being conducted.

In order to investigate effect of listener`s age, perception data for 7-year-olds are currently collected. While the procedure is the same as the one used for the adults, the set of child stimuli only includes 3-year-olds, 7-year-olds (as peer group) and adults.

[1] Hardcastle, W. J., & Hewlett, N. (Eds.). (2006). *Coarticulation: Theory, data and techniques*. Cambridge University Press.

[2] Beddor, P. S., McGowan, K. B., Boland, J. E., Coetzee, A. W., & Brasher, A. (2013). The time course of perception of coarticulation. *The Journal of the Acoustical Society of America, 133*(4), 2350–2366.

[3] Nittrouer, S., & Whalen, D. H. (1989). The perceptual effects of child–adult differences in fricative-vowel coarticulation. *The Journal of the Acoustical Society of America, 86*(4), 1266–1276.

[4] Whalen, D. H. (1983). Vowel information in postvocalic fricative noises. *Language and speech, 26*(1), 91–100.

[5] Bertoncini, J., & de Boysson-Bardies, B. (2000). Chapitre 4. La perception et la production de la parole avant deux ans. In *L'acquisition du langage. Vol. I* (pp. 95–136). Presses Universitaires de France.

[6] Houston, D. M., & Jusczyk, P. W. (2003). Infants' long-term memory for the sound patterns of words and voices. *Journal of Experimental Psychology: Human Perception and Performance, 29*(6), 1143.

[7] Noiray, A., Abakarova, D., Rubertus, E., Krüger, S., & Tiede, M. (2018). How do children organize their speech in the first years of life? Insight from ultrasound imaging. Journal of Speech, Language, and Hearing Research, 1–14.

[8] Rubertus, E., & Noiray, A. (2018). On the development of gestural organization: A cross-sectional study of vowel-to-vowel anticipatory coarticulation. *PloS one, 13*(9), e0203562.

[9] Sussman, H. M., Hoemeke, K. A., & McCaffrey, H. A. (1992). Locus equations as an index of coarticulation for place of articulation distinctions in children. *Journal of Speech, Language, and Hearing Research, 35*(4), 769–781.

[10] Nittrouer, S., & Studdert-Kennedy, M. (1987). The role of coarticulatory effects in the perception of fricatives by children and adults. *Journal of Speech, Language, and Hearing Research, 30*(3), 319–329.

[11] Sereno, J. A., Baum, S. R., Marean, G. C., & Lieberman, P. (1987). Acoustic analyses and perceptual data on anticipatory labial coarticulation in adults and children. *The Journal of the Acoustical Society of America, 81*(2), 512–519.

[12] Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & psychophysics, 28*(4), 267–283.

[13] Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44*(2), 314–324. doi:10.3758/s13428-011-0168-7.

[14] Noiray, A., Wieling, M., Abakarova, D., Rubertus, E., & Tiede, M. (2019, forthcoming). Back from the future: nonlinear anticipation in adults and children´s speech. *Journal of Speech, Language, and Hearing Research.*

Perception of Canadian English intervocalic and isolated sibilants: Processing of acoustic information or underlying (articulatory) vocal tract configurations?

San-Hei Kenny Luk[1] and Daniel Pape[1]
[1]*McMaster University, Hamilton, Canada*

Acoustic and articulatory speech perception theories aim to explain how listeners map the incoming acoustic signals to phonetic categories. In contrast to acoustic theories, articulatory theories (e.g. Motor Theory [1] or Direct Realist Theory [2]) hypothesize that listeners would recover the underlying articulatory information or motor commands, whereas acoustic theories only rely on the spectral content of the acoustic signal for phoneme identification. To test the hypothesis whether articulatory information can be recovered in isolated and intervocalic sibilant perception, we examined how manipulating different frequencies ranges will affect identification of Canadian English sibilants /s ʃ/. Specifically, we either (a) manipulated the amplitude of the frequency range where the main spectral peaks are found (i.e. around 3kHz for /ʃ/ and 7kHz for /s/) and, therefore, including front cavity and thus place of articulation information (from here on defined as *relevant* frequencies) or (b) the amplitude of frequency ranges where the sibilant place of articulation is *not* coded, i.e. articulatory irrelevant frequency ranges for the perception of a given sibilant (from here on defined as *irrelevant* frequencies with respect to front cavity resonance and thus place of articulation). The more interesting manipulation is altering *irrelevant* frequencies in the acoustic signal: While maintaining the underlying articulatory information, the manipulated sibilants would still be *articulatorily* cued as the original sibilants (i.e. defined by front cavity properties), however they would be *acoustically* cued as the alternative sibilants (i.e. /s/ as /ʃ/, and /ʃ/ as /s/), with completely changed acoustic spectral shape due to the strong acoustic manipulation (see figure 1).

We conducted an identification (speech perception) experiment with 31 listeners to examine how strong manipulation of acoustic information would switch Canadian English listeners' identification scores. Manipulated fricatives were either presented in isolation, re-embedded (after manipulation) in their original /a_a/ context or cross-spliced into their alternative sibilant's vowel context (e.g. fricative /ʃ/ embedded in /a_a/ context where /s/ was removed and replaced with /ʃ/). We found that the listeners' switch in identification categories strongly depends on the phoneme category of the *underlying* sibilant (see figure 2 left): Listeners identified acoustically /s/-like /ʃ/ completely as the alternative sibilant /s/ (see (1)), however the acoustically /ʃ/-like /s/ was perceived only at chance level as the alternative sibilant /ʃ/ (see (2)). In other words, manipulations of the *irrelevant* parts of the spectra with respect to the front cavity frequency range lead to a categorical switch for the *underlying* /ʃ/ stimuli but *not* for the *underlying* /s/ stimuli. The presentation of manipulated fricatives in original vowel context did not change these results. However, the cross-splicing in alternative sibilant vowel context ensured the previously lacking phoneme boundary shift for underlying /s/ stimuli (see figure 2 right), so the inclusion of the original /s/ formant transitions shifted listener responses towards 100% /ʃ/ perception. It seems that perceptual cue-trading between fricative noise and vowel context takes place here.

In conclusion, we here provide additional evidence that articulatory information can in fact be recovered for sibilants, but recovery strongly depends on the *underlying* phonetic category (i.e. in this case restricted to alveolars), possibly due to rivaling articulatory gestures like the presence (/ʃ/) or absence (/s/) of lip rounding gestures facilitating or hindering articulatory parsing. However, we also found that acoustic information clearly dominates articulatory information in the identification process, at least for the tested Canadian English voiceless sibilants.

(1) Manipulation of underlying /ʃ/ stimuli: Strong acoustic manipulation of the *irrelevant* frequency regions (no manipulation for the vocalic part) lead to a complete switch to the /s/ category, with stronger amplitude manipulations leading to more switches. Manipulating *relevant* frequency regions (steps 1 and 2) resulted in the expected ceiling effect of /ʃ/ judgments. The original recorded /ʃ/ stimulus is step 3.

(2) Manipulation of underlying /s/ stimuli: Strong acoustic manipulation of the irrelevant frequency regions did *not* lead to a complete switch towards the /ʃ/ category, with the highest manipulation of step 7 resulting in a mere chance judgment (around 50% /s/ responses and 50% /ʃ/ responses). Manipulating relevant frequency regions (steps 1 and 2) resulted in an expected ceiling effect of /s/ judgments. The original recorded /s/ stimulus is step 3.

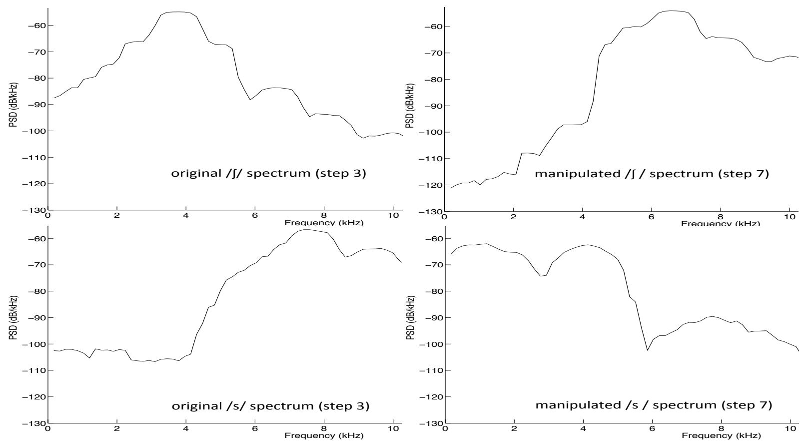


Figure 1. *Comparison between the acoustic spectra of the recorded stimuli (/s/ and /ʃ/, left panels) and acoustic spectra of the manipulated stimuli with strongly amplified irrelevant frequency range (right panels, step7 in figure 2 with +48dB amplification). As a result, the acoustically manipulated stimuli were acoustically very similar in overall spectral shape to the other examined sibilant (/s/ as /ʃ/ and /ʃ/ as /s/).*

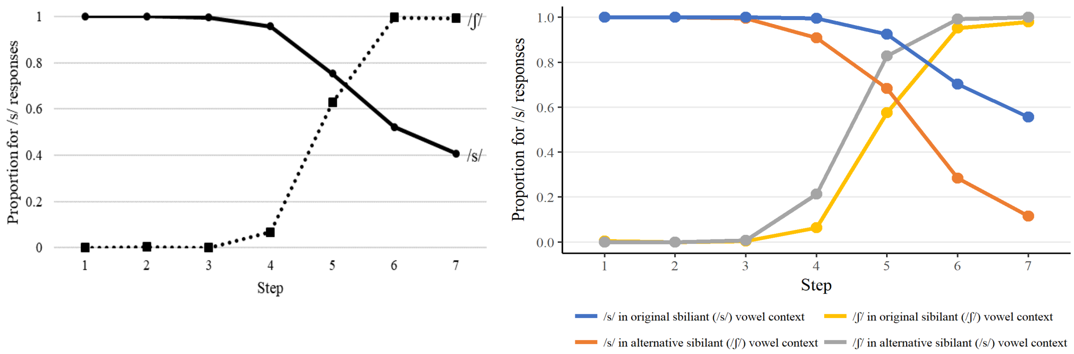


Figure 2. *Identification results for Canadian English sibilant perception (left: isolated fricative noises; right: fricatives in original vowel context or cross-spliced alternative sibilant vowel context. The original recorded sibilant (/s/ or /ʃ/) is step 3, amplified relevant frequency stimuli are steps 1+2, and amplification of irrelevant frequency regions are steps 4-7. The magnitude of the acoustic manipulation increases linearly from step 4 to step 7 (step4 = 12dB, step7 = 48dB). See text for definition of relevant and irrelevant frequency regions.*

[1] Galuntucci, B., Fowler C.A. & Turvey M.T. 2006 The motor theory of speech perception reviewed, *Psychonomic Bulletin & Review* 13(3), 361-377.

[2] Fowler, C.A. (1986). An event approach of the study of speech perception from a direct- realist perspective. *Journal of Phonetics*, 14, 3-28.

[3] Stevens, K. 1998. *Acoustic Phonetics.* MIT Press.

[4] Jongman A, Wayland R. & Wong S. (2000). Acoustic characteristics of English fricatives, *Journal of the Acoustical Society of America* 108, 1252-1263.

Poster session 4

Do Italian native learners expect geminate consonants in L2 French?

Paolo Mairano[1], Fabian Santiago[2]
[1]STL UMR 8163, Université de Lille (France); [2]SFL, Université de Paris 8 (France)

Introduction. A growing body of research shows that orthography influences the pronunciation and the phonological awareness of L2 speakers. The effects of orthography may go as far as to induce phonological oppositions that are inexistent in the target language: Bassetti et al. (2018) have shown that Italian native speakers produce longer consonants (geminates) in L2 English when spelled with one vs two consonants (e.g. *finish-Finnish*, realized as [ˈfɪnɪʃ]-[ˈfɪnːɪʃ]), replicating grapheme-phoneme conversion rules of their L1. The present study is inspired by such work but focuses on Italian learners of L2 French.

Consonant duration is phonological in Italian (incl. northern varieties of Italian, cf. Mairano & De Iacovo, submitted), but not in French. Mairano et al. (2018) replicated results by Bassetti and colleagues, showing that Italian learners of L2 French produce longer consonants in correspondence of <CC> spelling (e.g. *pratiquer-attirer*, *étape-échappe*). In this study, we investigate whether such learners *expect* longer consonants in correspondence of <CC> spelling. In order to test this hypothesis, we conducted a perception tests with 20 Italian learners at the University of Turin (B1 or above) and 20 native control speakers at the University of Paris 8.

Perception test. A cross-modal masked identification task was developed with *PychoPy* (Peirce, 2007). Audio stimuli consisted of French words uttered in isolation by a Parisian speaker with declarative intonation. All stimuli were manipulated in two ways: (i) they were partially masked by deleting either the first or the last syllable, and (ii) a target consonant was lengthened at 5 different conditions (no lengthening, +30%, +60%, +90, +120%), inspired by Rochet & Rochet (1995). We carefully selected 80 French words as stimuli with target consonants [p], [t], [l], or [m]/[n]: half of them were spelled as C (e.g. *signaler*), the other half as CC (e.g. *installer*); half were at the onset of the second syll. (e.g. *retirer*), half at the onset of the last syll. (e.g. *hôpital*). Participants heard a stimulus, then saw a word on the screen and had to decide whether the audio stimulus corresponded to the word on the screen. For instance: they heard [sjonal] (manipulated with a +60% longer [n]), then saw the word *national* and had to decide whether what they heard is a possible end for *national*. 40 distractors had a real incoherence between the audio and the word on the screen (e.g. [topys] instead of [tobys] for *autobus*). We adopted a Latin Square design so that each participant would never hear the same word more than once. We expected lower accuracy and longer reaction time for stimuli with mismatching cues (i.e. long <C> and short <CC>).

Results and discussion. The results (fig. 1) suggest that Italian learners recognise <CC> words equally if pronounced with a short or a long consonant. However, accuracy decreases significantly for <C> words lengthened by 90% and 120%. This effect is smaller than expected, but the difference with respect to native speakers (who show no effect) is visible.

Reaction times (fig. 2) for Italian learners present the expected X pattern with higher values for mismatching cues and lower values for matching cues, although not all differences reach statistical significance. Again, we observe a difference with respect to native participants, for whom consonant lengthening seems to have no effect whatsoever.

The production results reported by Mairano et al. (2018) showed that Italian learners of L2 French replicated grapheme-to-phoneme conversion rules of their L1 and thereby produced longer consonants in correspondence of <CC> than <C> spelling with a ratio of 1.2 (i.e. smaller than in their L1, as already found by Bassetti et al., 2018). This happens even in word-final position, where Italian phonotactics does not license geminate consonants. It is therefore argued that the orthographic effect under scrutiny affects more strongly L2 production than L2 perception.

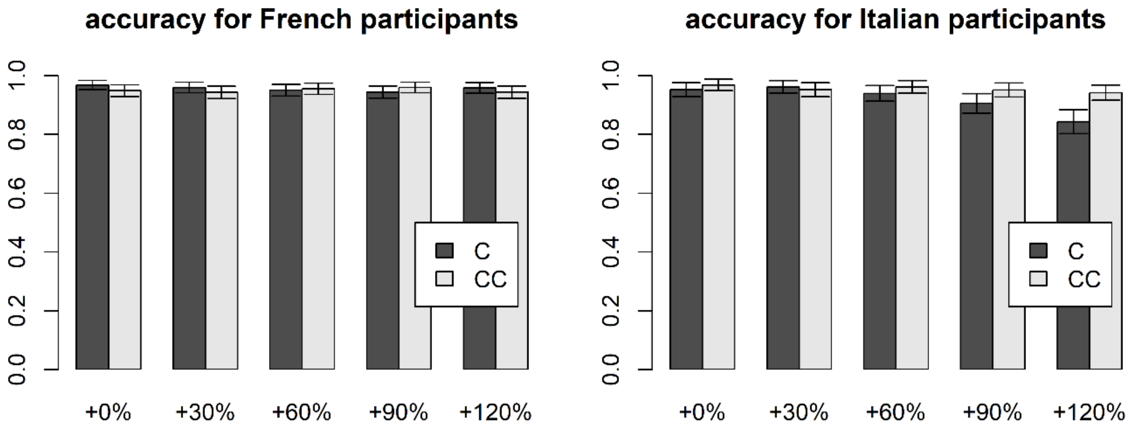


Figure 1. Accuracy (percentage of identified stimuli) in the 5 lengthening conditions for C vs CC words.

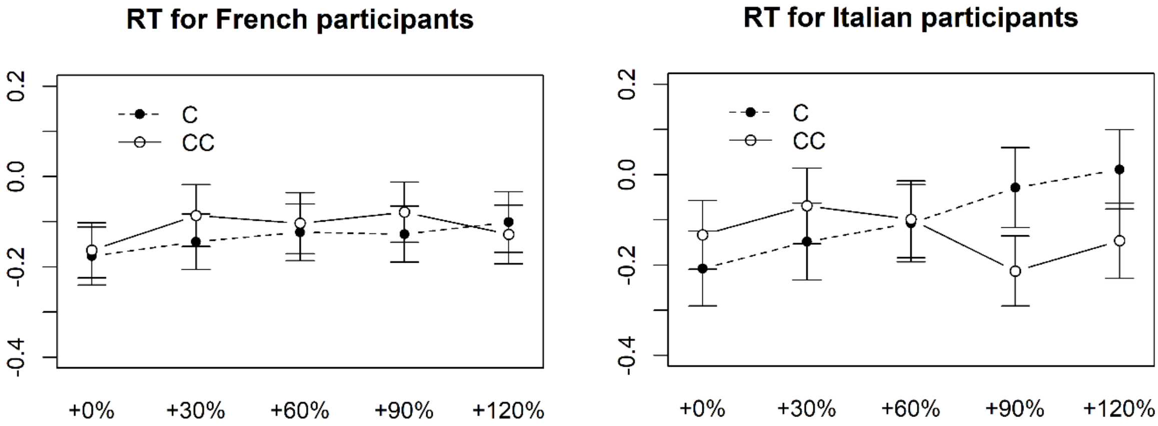


Figure 2. Z-score transformed reaction times for the identification of stimuli in the 5 lengthening conditions for C vs CC words.

References

Bassetti, B., Sokolovic-Perovic, M., Mairano, P. & Cerni, T. (2018) Orthography-induced length contrasts in the second language phonological systems of experienced speakers of English as a Second Language: evidence from minimal pairs. *Language and Speech*, 61.4, 577-598.

Mairano, P. & De Iacovo (submitted) Gemination in northern vs central and southern varieties of Italian: a corpus-based investigation.

Mairano, P., Santiago, F. & Delais-Roussarie, E. (2018) Gémination non-native en français d'apprenants italophones. *Actes des Journées d'Etude sur la Parole (JEP)* (pp. 657-665), Aix-en-Provence (France), 4-8 juin 2018.

Peirce, J.W. (2007). PsychoPy-psychophysics software in Python. *Journal of Neuroscience Methods*, 162(1-2), 8-13.

Rochet, L.B., & Rochet, A.P. (1995). The perception of the single-geminate consonant contrast by native speakers of Italian and Anglophones. *Proc. of ICPhS95*, 616-619.

Accomodation processes in the speech of University College Students:

a real-time sociophonetic approach

Nicholas Nese

University of Pavia

The purpose of this research is to identify and quantify language variation in real time in a community of students in the university college “Giasone del Maino” in Pavia (Italy). We chose this sociolinguistic setting firstly for the heterogeneity of the student population: as students come from different regions of Italy, the college setting provides an optimal situation of language contact between different Italian regional varieties. Moreover, the peculiarity of collegial life allows us to consider the student body as one *community of practice* (Eckert and McConnell-Ginet:1992; Wenger:1998), which, in our hypothesis, would accelerate the processes of diachronic variation. Eckert also notes that language change in students is related in particular to the practices associated with their own group and not to the socio-economic class of their family. This study, in parallel with the work of Nodari (2017), is one of the first research that focus on the Italian language and based on the Eckert-like model.

Data were collected through interviews aimed at eliciting three distinctive speech styles at different level of control over the speech production: spontaneous speech, through a short self-presentation of the speaker and some general questions; semi-spontaneous speech, through different linguistic tasks; and controlled speech, through the reading of a word list. The whole corpus has been designed to investigate different phonemes related to both vocalism and consonantism of Italian. The word list reading task includes 49 tokens of real words. To study diachronic variation, the data were collected at different points throughout the academic year.

For this poster-presentation, a case-study of 6 speakers (3 male and 3 female) of two different linguistic areas (3 from northern Italy and 3 from southern Italy) have been analyzed. Data were collected at the freshmens’ arrival in college (Freshmen T1) and at 11 months after arrival (Freshmen T2). A first analysis was performed on dental affricates in the word list reading (144 tokens). As demonstrated by previous analysis (see Meluzzi 2016), these phonemes present three sonority degrees (voiced, voiceless and mixed) with a high degree of variation across the Italian regional varieties. A comparison of the data relating to the sonority degree of Freshmen T1 shows a prevalence of voiced outcomes within southern students (32.4% voiceless, 17.6% mixed and 50% voiced) with respect to the northern students (52.8% voiceless, 30.6% mixed and 16.7% voiced) [$p=.012$]. After only 11 months, it is possible to see in Freshmen T2 an important reduction of the voiced dental affricates of the students of the south (50% voiceless, 13.9% mixed and 36.1% voiced) which are close to the outcomes of the students of the north (58.3% voiceless, 19.4% mixed, and 22.2% voiced) [$p>.05$]. In the Freshmen T2 data, the difference of north-south origin is no longer statistically significant (Pearson's chi-squared test).

These first results, although partial, suggest the presence of a process of accommodation towards the northern Italian regional variety by southern speakers. This phenomenon can be justified both by the perception of a greater prestige of the northern Italian regional variety and by the influence of the peculiarity of the community of practice.

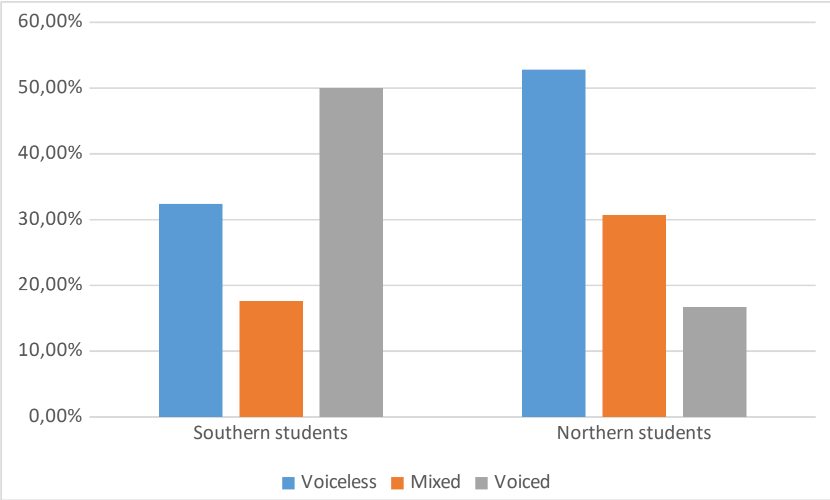


Figure 1. Freshmen T1 dental affricates [p=.012]

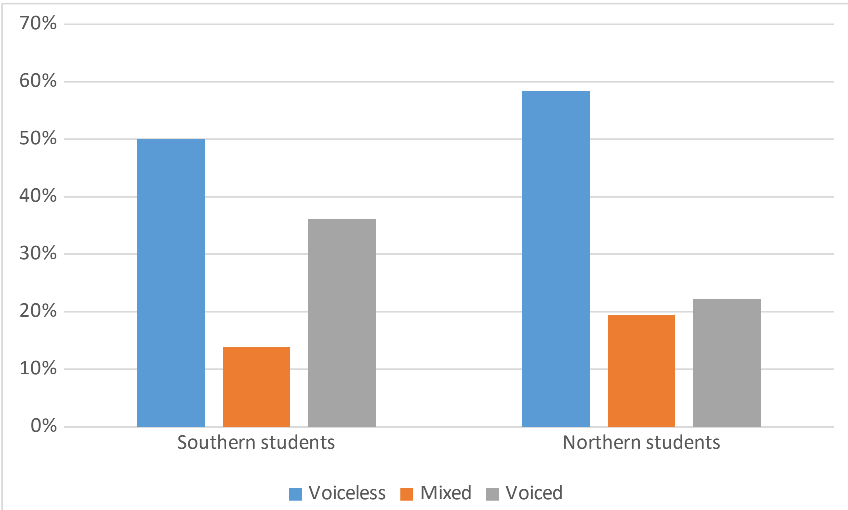


Figure 2. Freshmen T2 dental affricates [p>.05]

[1] Eckert, P. & McConnel-Ginet S. 1992. *Communities of practice: Where language, gender and power all live.* In Kira Hall, Mary Bucholtz and Birch Moonwomon eds. *Locating Power,* 89-99.

[2] Meluzzi, C. 2016 *A new sonority degree in the realization of the dental affricates /ts dz/ in Italian*, in M.J. Ball & N. Muller (eds.) Challenging Sonority: Cross-Linguistic Evidence from Normal and Disordered Language, London: Equinox, pp. 252- 275.

[3] Nodari, R. 2017 *L'italiano degli adolescenti: aspirazione delle occlusive sorde in Calabria e percezione della varietà locale,* Scuola Normale Superiore di Pisa: Tesi di dottorato in Linguistica.

[4] Van der Harst S. 2011 *Vowel Space Paradox. A Sociophonetic Study on Dutch*, Utrecht LOT

[5] Wenger, E. 1998. *Communities of practice: Learning, meaning and identify.* Cambridge, Cambridge University Press.

Variability in the perception of epistemic valence in Salerno Italian question tunes

Riccardo Orrico[1,2], Renata Savy[1], Mariapaola D'Imperio[2,3]
[1]*University of Salerno, Italy*
[2]*Aix Marseille University, CNRS, LPL UMR 7309, Aix en Provence, France*
[3]*Rutgers University, New Brunswick (NJ), USA*

Intonation can express a wide range of linguistic information, such as sentence modality, information structure, speaker attitudes, and epistemic valence. However, one issue that still remains controversial is the stability of the relationship between intonational form and meaning, which suggests the absence of variability in the way intonation is used across speakers to express meanings. This position has been challenged by [1], showing speaker-dependent strategies in the tonal signalling of focus types in German, and by [2], arguing the absence of a one-to-one correspondence between tunes and linguistic functions. Recent studies have tried to model variability in production, by assuming that it is a systematic rather than a random phenomenon, linked to speaker-specific features (gender, age, dialect) [3]. Such features might also affect perception and shape the way listeners make inferences about speakers' intentions. Nevertheless, one uncontroversial point is the role of intonation in coding information regarding speaker epistemic attitude, which has been argued for several languages (English [4], French [5], Catalan [6], and Bari Italian [7]).

Our study intends to investigate how variability of polar question tunes is employed in the identification of epistemic modality in Salerno Italian (SI). The tunes used for SI polar questions are two phonologically distinct rise-falls, L*+H L-L% and L+H* L-L%, and a rise-fall-rise, L+H* L-H% (see Figure 1), whose distribution appears to be highly dependent on individual speakers [8]. The hypothesis we test is that such variability depends on a complex interplay between the expression of speaker commitment and individual variability in terms of chosen intonational form on the part of the listener. We also hypothesise that such variability might be explained by listeners' exposure to other languages or dialects (for example, whether they lived in cities other than the native one), assuming that phonological traces from past experience with other systems might affect the perception of the native variety [9].

We performed a perception survey (submitted through Gorilla.sc) and asked 45 SI listeners to judge question utterances (24 items x 3 tune types) according to perceived degree of speaker certainty relative to the question response. Certainty scores were collected using a slider ranging from 0 ('She expects no') to 100 ('She expects yes').

Data were analysed using mixed effect linear regression models in R [10], using the packages *lme4* [11] and *afex* [12].The model included Tune type (3 levels) and Other cities (having lived or not in another city for more than 12 months, 2 levels) as fixed effects. We also included Listeners and Items as random intercepts and slopes for the effect of Tune type.

Results show an effect of Tune type, $F(2, 32.07) = 4.87$, $p < .01$. Specifically, multiple comparisons among Tune type levels revealed that L*+H L-L% received lower certainty values, however the only pair of tunes that yielded significantly different results was L*+H L-L% vs. L+H* L-H%, while the inspection of the other two contrasts revealed no significant difference. Also, an effect of Other cities was found ($F(1, 47.32) = 5.89$, $p < .02$): listeners who lived in other cities globally rated all stimuli with a lower certainty value relative to all other listeners. Additionally, random intercepts for Listeners show that individual differences in response scores appear to be quite important (see Fig. 2, right).

As hypothesized, the significant effect of Tune type on the Certainty score show that epistemic valence can be tonally encoded in SI question tunes. On the other hand, listener's behaviour was influenced by sociophonetic (Other Cities) as well as by additional idiosyncratic factors, suggesting that the link between intonational form and pragmatic meaning cannot be uniformly modelled within a language community.

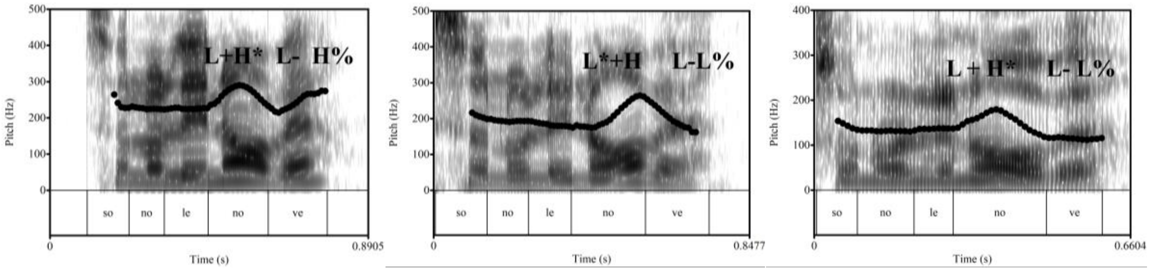


Figure 1. *F0 contours for the question Sono le nove? "Is it nine o'clock?" uttered with either the L*+H L-L% (left), L+H* L-H% (middle), or the L+H* L-L% (right) tune.*

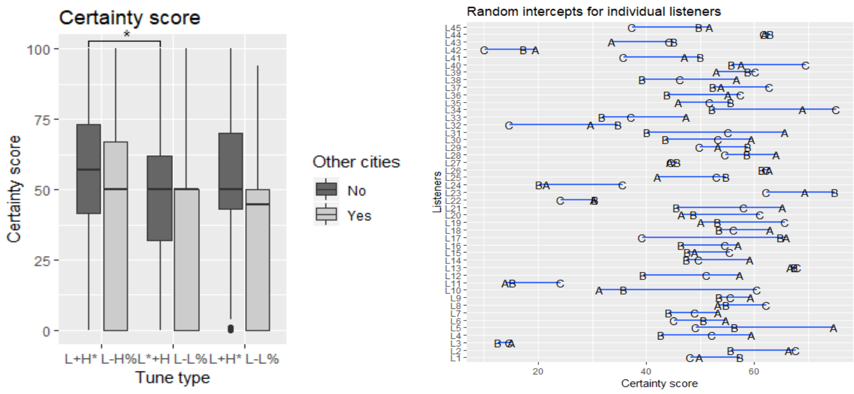


Figure 2. *Response scores by Tune Type and Other Cities (left) and Individual Listeners intercept values, A: L+H* L-H%, B: L*+H L-L%, C: L+H* L-L% (right).*

References

[1] Grice, M., Ritter, S., Niemann, H., & Roettger, T. B. (2017). Integrating the discreteness and continuity of intonational categories. *Journal of Phonetics*, 64, 90-107.

[2] Gili Fivela, B. (2008). Intonation in production and perception: The case of Pisa Italian. Edizioni dell'Orso.

[3] Kleinschmidt, D. F. (2019). Structure in talker variability: How much is there and how much can it help? *Language, cognition and neuroscience*, 34(1), 43-68.

[4] Gunlogson, C. (2004). True to form: Rising and falling declaratives as questions in English. Routledge.

[5] Portes C., Beyssade C. (2012). Is intonational meaning compositional? *Verbum*, 34.

[6] Prieto, P. & Borràs-Comes, J. (2018). Question intonation contours as dynamic epistemic operators. *Natural Language and Linguistic Theory*, 36 (2), 563-586.

[7] Savino, M., & Grice, M. (2011). The perception of negative bias in Bari Italian questions. In Frota, S. Prieto, P., Elordieta, G. (Ed.), *Prosodic categories: Production, perception and comprehension*. Springer: Dordrecht, 187-206.

[8] Orrico, R., Savy, R., & D'Imperio, M. (submitted). XV Convegno AISV, Dipartimento di Scienze della Formazione, Scienze umane e della Comunicazione interculturale, Arezzo, 14-16 February 2019.

[9] Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological review*, 105(2), 251.

[10] R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

[11] Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version, 1(7), 1-23.

[12] Singmann, H., Bolker, B. Westfall, J. & Aust, F. (2018). afex: Analysis of Factorial Experiments. R package version 0.22-1. https://CRAN.R-project.org/package=afex

Rhythmic Convergence and Divergence in two Swiss German dialects

Elisa Pellegrino[1,2], Hanna Ruch[1], Thayabaran Kathiresan[2], Volker Dellwo[2]
[1]*URPP Language and Space, University of Zurich,*
[2] *Department of Computational Linguistics, University of Zurich*

A sizeable body of research has showed that during social interactions speakers accommodate to their interlocutors, either by becoming more similar (convergence) or by accentuating individual differences (divergence) [1]. With respect to existing evidence of rhythmic adjustments in response to the interlocutor's age and cognitive development [2], in the present study we test the hypothesis that rhythmic modifications are not only unidirectional but that interlocutors are likely to mutually adapt their rhythmic characteristics over the course of a conversation or after increased exposure to a dialogue partner. One excellent testing ground for studying rhythmic accommodation is represented by the linguistic situation of German-speaking Switzerland. Swiss German dialects, differ for segmental features, speech rate and intonation contours [3-7] as well as for their rhythmic characteristics, measured acoustically in terms of the durational variability of consonantal and vocalic intervals [8]. Particularly, the two Swiss German dialects under study (Grison and Zurich German) present three durational crucial differences, leading to a different rhythmic organisation of the two dialects. Compared to Zurich German (ZH), Grison German (GR) presents: opens syllable lengthening (OSL), intervocalic sonorants gemination (ISL) and unreduced word-final ending (Red@) (cf. Table 1).

To study rhythmic accommodation in a dialect contact situation, we used a corpus composed of: 18 audio-recorded dialogues between ZH and GR speakers while performing a diapix task, and 18 pre- and 18 post-dialogue recordings (picture naming task and retelling a story based on a comic), these two latter performed individually by ZH and GR participants [9]. To determine whether ZH and GR speakers produce the rhythmic contrasts more similarly after participating in the diapix task, we we extracted the lexical items instantiating the three target durational contrasts from the pre- and post dialogue recordings; in pre-dialogue recordings we measured the cross dialectal differences in three ratio measures devised to capture the rhythmic differences between the two dialects (OSL: ratio between stressed and unstressed vowel within the same word; - ISG: ratio between intervocalic sonorants in -CCe words and in -Ce words;- Red@: the ratio between word-final ending and stressed vowel) (fig. 2) and then we measured difference in distance within a pair (*ddpair*) in the three ratio measures before and after the interaction.

Results based on picture naming task showed that, despite the significant cross-dialectal differences in the three durational contrasts in pre-dialogue session [ISG: $t(34) = -7.816$, $p < 0.001$; OSL: t(test) (34) = $p < 0.001$; Red@: $t(34) = 3.161$, $p = 0.03$], pairs do not shift consistently the production of the three durational contrasts after the dialogue. Visually, some pairs seem to converge more than others, but the results of statistical analysis (one-way ANOVA, with "Pair" as Factor) report no significant cross-pair differences in the three durational characteristics (*ddpair* in ISG: $F(17, 72) = 1{,}52$, $p = 0.119$; *ddpair* in OSL: $F(17, 126) = 0.793$, $p = 0.698$; *ddpair* in Red@: $F(17, 252) = 1.480$, $p = 0.102$). In light of Trudgill's assumption that short term accommodation may bring about language change [10], the results of this study will shed light on which acoustic features of the two dialects play a role in dialectal levelling and diffusion of linguistic innovations. Although these findings cannot be over-interpreted, as they are based only on words pronounced in isolation, they can be taken as an indication that durational contrast, unlike vowel quality, are probably too subtle be perceived and thus to be imitated.

Dialect feature	Example with translation	GR		ZH	
OSL	Sohle ‘sole’	V:	[ˈsoːlɐ]	V	[ˈsolə]
ISG	Pille ‘pill’/ Sonne ‘sun’	ll/nn	[ˈpillɐ]/[ˈsunnɐ],	l/n	[ˈpilə]/ [ˈsunə]
Red@	Suppe ‘soup’	ɐ	[ˈsuppɐ]	ə	[ˈsuppə]

Table 1. Examples of items in GR and ZH for the three durational contrasts (adapted from [11])

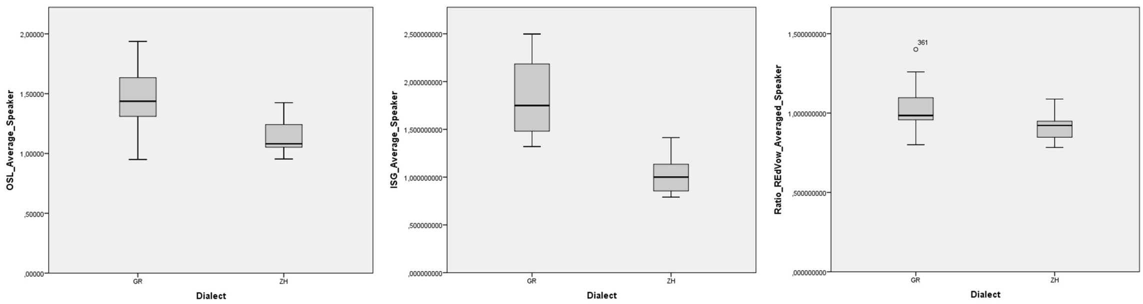


Figure 1. Cross dialectal differences in three durational contrasts averaged per speaker (left: ISG, centre: OSL; right: Red@)

References

[1] Giles, H. & Ogay, T. 2007. Communication Accommodation Theory. In B. B. Whaley & W. Samter (Eds.), Explaining communication: Contemporary theories and exemplars (pp. 293-310). Mahwah, NJ: Lawrence Erlbaum

[2] Payne, E., Post, B., Prieto, P., Astruc, L. & Vanrell, M.d. M. 2010. Rhythmic Modification of Child Directed Speech. In M. Russo (Eds): Prosodic Universals. Comparative Studies in Rhythmic Modeling and Rhythm Typology, Roma: Aracne Editrice, 147-183.

[3] Eckhardt, O. 1991. Die Mundart der Stadt Chur. Zürich: Phonogrammarchiv der Universität 624 Zürich.

[4] Christen, H., Glaser, E., and Friedli, M. 2010. Kleiner Sprachatlas der deutschen Schweiz. Frauenfeld: Huber Frauenfeld.

[5] Fleischer, J., & Schmid, S. 2006. Zurich German. J. Int. Phon. Assoc. 36, 243–253.

[6] Ruch, H. 2015. Re-emergence of sonorant geminates in Zurich and Grison German. Poster 702 presented at Geminate Consonants across the World, Satellite Workshop at the ICPhS 2015, 12 August 2015, Glasgow.

[7] Leemann, A. 2012. Swiss German Intonation Patterns. Amsterdam; Philadelphia: John Benjamins. Publishing.

[8] Leemann, A, Dellwo, V.; Kolly, M.-J., Schmid, S. 2012. Rhythmic variability in Swiss German dialects. In: 6th International Conference on Speech Prosody, Shanghai, 22 May 2012 - 25 May 2012, 607-610.

[9] Ruch, H. 2015. Vowel convergence and divergence between two Swiss German dialects. In: 18th International Congress of Phonetic Sciences, Glasgow UK, 10 August 2015 - 14 August 2015.

[10] Trudgill, P. 1986. Dialects in Contact. Oxford: Blackwell Publishing.

[11] Ruch, H. (2018). The role of acoustic distance and sociolinguistic knowledge in dialect identification. Frontiers in Psychology.

The speech production system is reconfigured to change speaking rate

Joe Rodd[1,2] Hans Rutger Bosker[1,3] Mirjam Ernestus[2,1] Antje S. Meyer[1,3] Louis ten Bosch[2,1]
[1]*Max Planck Institute for Psycholinguistics,* [2]*Radboud University, Centre for Language Studies,* [3]*Radboud University, Donders Institute for Brain, Cognition and Behaviour*

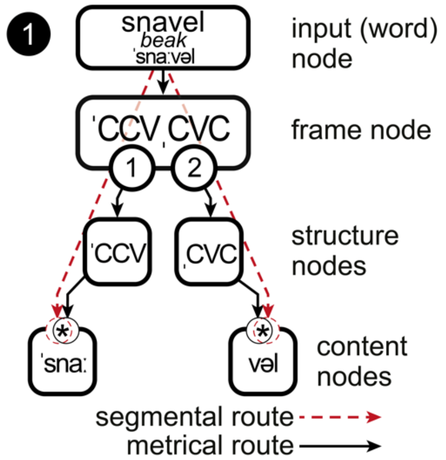
It is evident that speakers can freely vary stylistic features of their speech, such as speech rate, but how they accomplish this has hardly been studied, let alone implemented in a formal model of speech production. Much as in walking and running, where qualitatively different gaits are required cover the range of different speeds, we may predict there to be multiple qualitatively distinct configurations, or 'gaits', in the speech planning system that speakers must switch between to alter their speaking rate or style. Alternatively, control might rely on continuous modulation of a single 'gait'. This study investigates these possibilities through simulations of a novel connectionist computational model of the cognitive process of speech production, which mimics the temporal characteristics of observed speech.

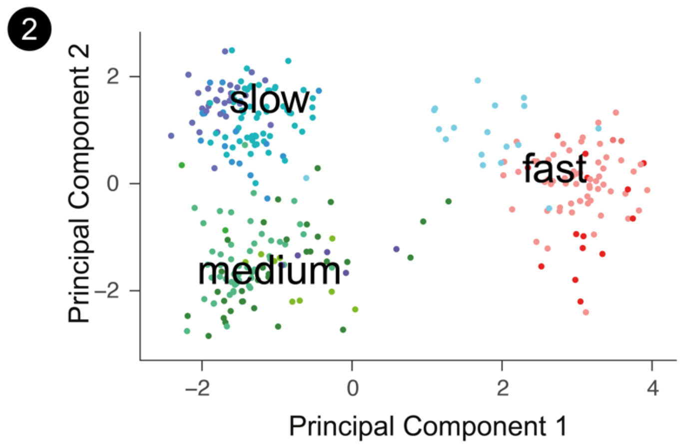
Connectionist model Our model (Figure 1) is derived from Dell, Burger and Svec's [1] model of serial order in language production, and sequentially retrieves syllable-level motor plans in response to activation in a word level input node. A frame node mediates, encoding metrical structure and enforcing serial order. This model is the first of its type to predict the precise timing of motor plans and account for the ability to control rate in speech production. The model has many parameters (connection weights, thresholds, etc.) that can be adjusted to fit a speaking rate. Different 'regimes' (combinations of parameter settings) can be engaged to achieve different speaking rates. We consider each parameter as a dimension of a high-dimensional 'regime space', in which the regimes occupy different locations.

Model training Our model approximated the distributions of observed syllable durations and syllable overlap durations in the PiNCeR corpus of Dutch disyllabic words produced at fast, medium and slow speaking rates. Syllable onset and offset were identified from the acoustic signal on the basis of spectral instability as an index of syllable overlap. Together, these duration distributions form a 'fingerprint' of the speech production system operating at a given rate. The model was trained separately for each speaking rate, by the evolutionary optimisation algorithm NSGA-III [2]. The training identified parameter values that resulted in the model to best approximate the duration distributions characteristic of each speaking rate. The fit of the model was assessed by calculating the Kullback-Leibler divergence between the model's predicted distributions and those taken from the corpus for each speaking rate.

Predictions In one gait system, where we 'speed-walk' to speak faster, the regimes used to achieve fast and slow speech are qualitatively similar, but quantitatively different. In regime space, they would be arranged along a straight line. Different points along this axis correspond to different speaking rates. In a multiple gait system, where we 'walk-speak' for slower speaking rates, but 'run-speak' to speak faster, this linearity would be missing. Instead, the arrangement of the regimes would be triangular, with no obvious relationship between the regions associated with each gait, and an abrupt shift in parameter values to move from speeds associated with 'walk-speaking' to 'run-speaking'.

Results Our model achieved good fits in all three speaking rates. In regime space, the arrangement of the parameter settings selected for the different speaking rates is clearly not triangular, suggesting that 'gaits' are present in the speech planning system (Figure 2). Further models fitted at intermediate points in regime space between the speaking rates revealed stark non-linearities between slow and medium and between slow and fast, but not between medium and fast (Figure 3). This leads us to conclude that one configuration is engaged for medium and fast speech, and a second qualitatively distinct configuration is engaged for slow speech. Thus, we provide the first computationally explicit connectionist account of the ability to modulate the speech production system to achieve different speaking styles.





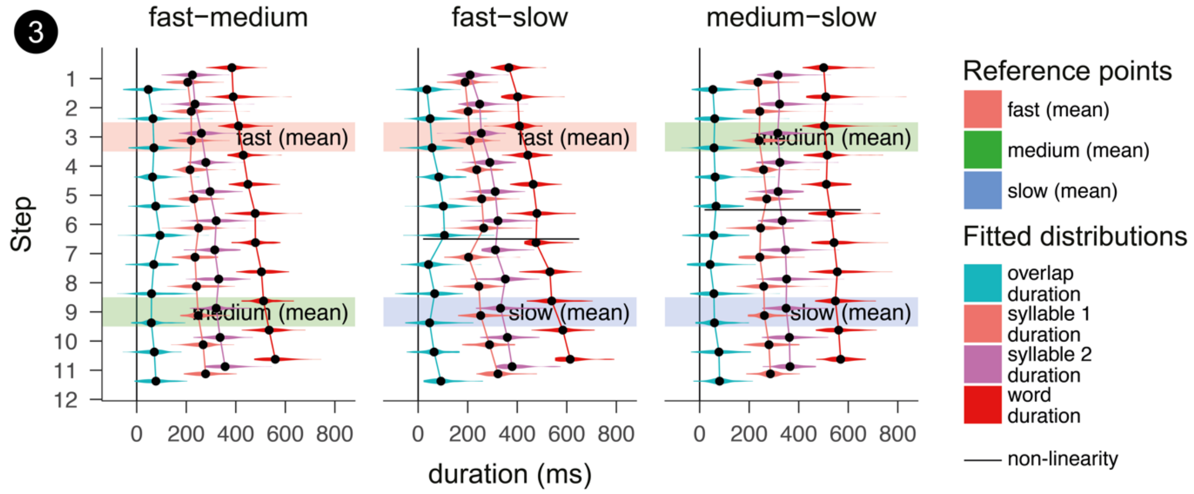


Figure 1. *The connectionist model, showing the nodes and connections required to produce the two syllables of the disyllabic Dutch word "snavel".*

Figure 2. *The best fitting parameter values for the three rate conditions, projected into PC1-PC2 space, showing the triangular arrangement of the parameters that best fit each rate.*

Figure 3: the predicted 'fingerprint' syllable duration, overlap duration and word duration distributions of interpolated points in regime space, illustrating the non-linearities (horizontal black lines) between fast and slow and medium and slow, and the absence of a non-linearity between fast and medium.

[1] G. S. Dell, L. K. Burger, and W. R. Svec, "Language production and serial order: A functional analysis and a model." *Psychological review*, vol. 104, no. 1, pp. 123–147, 1997.

[2] K. Deb and H. Jain, "An Evolutionary Many-Objective Optimization Algorithm Using Reference-Point-Based Nondominated Sorting Approach, Part I: Solving Problems With Box Constraints," *IEEE Transactions on Evolutionary Computation*, vol. 18, no. 4, pp. 577–601, 2014.

Clause Typing and Edge Tones in Early Italian: a longitudinal study

Martina Sciuto, Vincenzo Moscati, Giuliano Bocci,
University of Siena

Background. Very young infants are able to exploit the prosodic and metric properties of their language to uncover important syntactic regularities and to increase their lexicon (a.o. Christophe et al 2003, de Carvahlo et al 2017). However, it does not follow that the whole inventory of intonational configurations is mastered from very early on (see Chen 2011): although some general form-function correlations could be universal (Gussenhoven 2001), children still need to associate language-specific intonational patterns characterizing the target language they are exposed to.
Early intonation has been recently analysed adopting the Autosegmental Metrical (AM) framework, laying down the foundations for both cross-linguistic and longitudinal comparisons (Prieto et al. 2011 for Catalan and Spanish; Chen & Fikkert 2007 for Dutch; Frota & Vigario 2008 for European Portuguese). In this scenario, Italian data are still scattered and do not offer themself to a straightforward comparison (see d'Orrico & Carubbi 2003, d'Orrico et al. 2009). To start filling this gap, we will present the results of a preliminary investigation using the AM framework, analysing a new corpus of spontaneous production by an Italian-speaking child. We aim at: i) determining if the final rise expected in Yes/No questions (Y/NQs) is present from the very early productions and ii) if it univocally identifies Y/NQs, distinguishing them from declaratives and also from Wh-questions.

The Study. We collected a new corpus of recorded audio files from the spontaneous speech of an Italian Speaking girl, *'A'* who lives in Florence. The corpus is a high-density one (Tomasello & Stahl 2004) (>5h/week recordings) and covers a time-window from 19 to 34 months. 3 samples were selected for the intonational analysis, covering a one-year span: from the onset of the 2-words stage (Period 1, month 22, 7 files) to the last available recordings (Period 3, month 34, 12 files), plus an intermediate stage (Period 2, months 28-29, 16 files). Files were transcribed in ELAN (Wittenburg et al. 2004) using the CHAT format (MacWhinney 2000). Based on the pragmatic contexts, sentences were tagged into Declaratives, Yes/NoQs, and Wh-Qs. We isolated thus 486 sentences: see Table 1. As for Period 1, all the Declaratives and Y/NQs were taken (Wh-Qs are unattested). As for Period 2 and Period 3, all Y/NQs and Wh-Qs were included, plus a random sample of Declaratives roughly equalling the total number of Y/NQs and Wh-Qs. This set of sentences was intonationally annotated with the ToBIt transcription. In relation to our research questions, we focus on *A's* early production of rising boundary tones (H%), which is a prototypical feature of Y/NQs in most varieties of Italian (see Gili Fivela et al. 2015).

Results. The results, plotted in Fig.1., show a clear distinction between declaratives, on the one hand, and Y/N and Wh-Qs, on the other. Declaratives are overwhelmingly characterized by a low final contour from Period 1 (Fig.2). Y/NQs – whose number becomes relevant from Period 2 – are predominantly realized with a final rise (Fig.3). Wh-Qs (Fig.4), from their first appearance in Period 2, strikingly pattern with Y/NQs.

Conclusions. Our study shows that from the onset of the 2-word stage, *A* assigns different boundary tones to Declaratives and Y/NQs, a trend that is longitudinally stable. Furthermore, our data do not show any distinction between Y/NQs, and Wh-Qs, since the final rise equally mark both types of interrogatives.
In the light of the literature, the presence of such a large number of H% in Wh-Qs is unexpected, since the final rise is reported to be at most optional for Wh-Qs in the different varieties of Italian (see Gili Fivela et al. 2015). This finding could receive at least two types of explanations: a Target-deviant explanation: *A.* overgeneralizes the Y/NQ final rise to Wh-Qs, associating the same contour to all interrogatives; a Target-consistent explanation: the final rise produced by *A.* simply reproduces a specific feature of child-directed speech (i.e. adults "*hyperarticulate"* questions with a final rise when speaking to children). Discerning between the two HPs is an empirical matter that can be addressed – in progress - by a further look at adult's Wh- productions in their child-directed speech.

Sentence Type	Period 1 (22 Months, MLU = 1.34)	Period 2 (28 Months, MLU = 2.46)	Period 3 (34 Months, MLU = 3.19)
Declarative	40	74	146
Yes/No	2	35	74
Wh-	0	39	76
Tot	**42**	**148**	**296**

Table 1. Classification of *A's* sentences in the three sample periods.

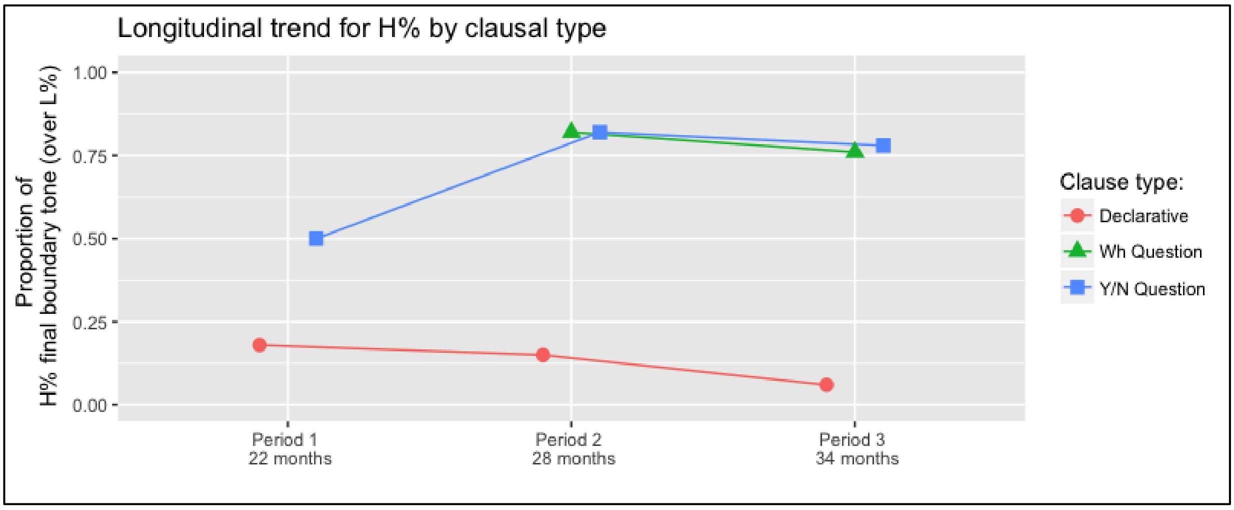


Figure 1. Proportion of *A*'s production of final rising tones. Longitudinal data by clause type.

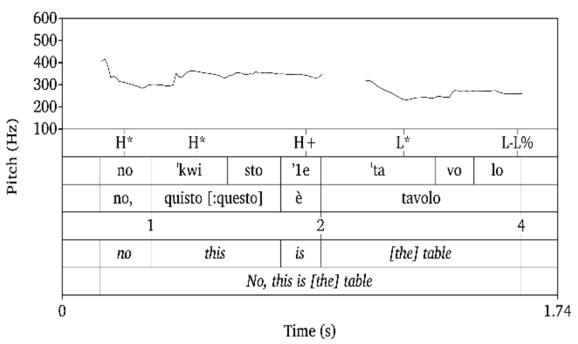


Fig.2 Ex. of a declarative with a L% boundary tone

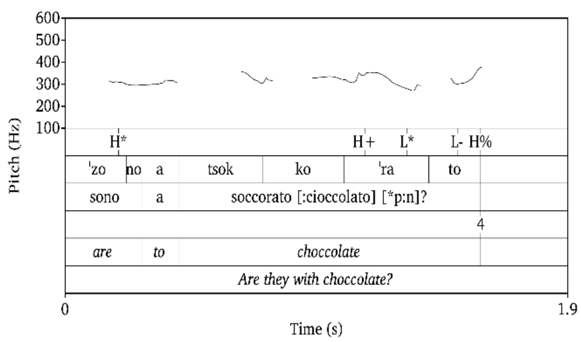


Fig.3 Ex. of a Y/NQ with a H% boundary tone

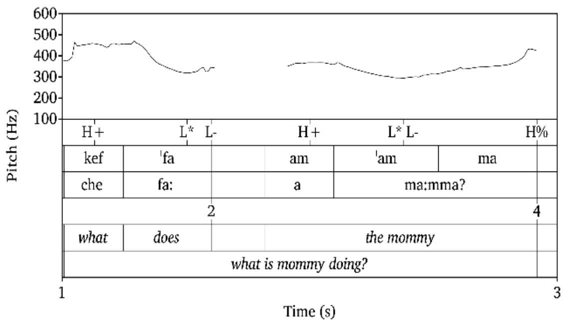


Fig.4 Example of a Wh-Q with a H% boundary tone

Selected references. DE CARVALHO, A., Dautriche, I., Lin, I., & Christophe, A. *(2017)*. Phrasal prosody constrains syntactic analysis in toddlers. *Cognition*, 163, 67-79. CHRISTOPHE, A., Gout, A., Peperkamp, S. & Morgan, J. (2003). Discovering words in the continuous speech stream: The role of prosody. *Journal of Phonetics* 31,585–98. D'ODORICO, L. and Carubbi, S. (2003). Prosodic Characteristics of Early Multi-Word Utterances in Italian Children. *First Language* (23).1, 97–116. GILI-FIVELA, *B.*, Avesani, C., et al.. (2015). Varieties of Italian and their intonational phonology. In Frota, S. & Prieto, P. (Eds.) *Intonational Variation in Romance*. Ch: 5,. Oxford: OUP. PRIETO, P., Estrella, A., Thorson, J. & Vanrell, M. M. (2012). Is prosodic development correlated with grammatical and lexical development? *Journal of Child Language* 39(2), 258–83.

How do French Cʁ# cluster realizations vary across speech styles?

Yaru WU[1,2], Martine ADDA-DECKER[2], Cécile FOUGERON[2], Cédric GENDROT[2]
[1]*Modèles, Dynamiques, Corpus (MoDyCo), UMR 7114, CNRS,*
[2]*Laboratoire de Phonétique et Phonologie (UMR7018, CNRS-Sorbonne Nouvelle), France*

The realization of consonant clusters tends to vary under the influence of the Sonority Sequencing Principle (SSP, [1][2][3]). French Cʁ# clusters are known to have several variants in order to avoid the violation of the SSP. Nonetheless, the distribution of the variants of Cʁ# clusters in spontaneous speech and the different strategies speakers apply across speech styles with regard to SSP violation are not fully understood. In this study, we will be concerned with obstruent + ʁ clusters, as in the word "quatre" (/katʁ/, *four*), which are the only Cʁ# clusters allowed in French. We investigate the realization of Oʁ# clusters when immediately followed by a word starting with a consonant (i.e. Oʁ#C). These 3-consonant sequences are challenging with respect to SSP (the SSP violation is resolved when the Oʁ# cluster is followed by a vowel). Besides the canonical realization (e.g. [katʁ̥]), speakers may drop the post-consonantal /ʁ/ (eg. [kat]), insert a schwa (eg. [katʁə]), or even delete /ʁ/ and insert schwa (eg. [katə]). The aim of this study is to better understand the distribution of these forms in continuous speech and how this distribution varies across different speech styles.

Three large French speech corpora were used for our investigations: ESTER (≈100 hours) [4], ETAPE (≈50 hours) [5] and the Nijmegen Corpus of Casual French (NCCFr, ≈40 hours) [6]. Both ESTER and ETAPE are broadcast speech. The ESTER corpus mainly contains formal journalistic news speech, whereas the ETAPE corpus is mostly composed of conversational journalistic speech, including conversations and debates. The NCCFr corpus contains casual conversations between friends.

Speech files were automatically aligned with the help of the LIMSI alignment system. For our study, we introduced specific pronunciation variants (i.e. /ʁ/ and schwa variants) allowing for optional /ʁ/ and schwa in all words ending in Oʁ# [7][8]. The aligned variant is considered reflecting the realized pronunciation. A subset of the automatic alignment (≈30h) is accompanied by a manual alignment carried out by an experienced French phonetician. Cohen's kappa [9] reveals that the two types of alignment have "almost perfect agreement" (kappa > 0.8).

The absence/presence of /ʁ/ and of schwa were determined by comparing the aligned pronunciations (reflecting the speakers' pronunciations) with the canonical pronunciation containing the 2 elements of the clusters without a schwa (eg. /katʁ/), as specified in Lexique380 [10]. A generalized linear mixed model (GLMM) [11] and follow-up post-hoc analyses were carried out to validate our analyses on the effect of speech style.

Figure 1 shows the overall results where variant production rates are pooled using the 3 corpora. The distribution of the variants of Oʁ# per speech style is shown in Figure 2. Overall, speakers tend to either insert schwa (47%) or delete /ʁ/ (33%) when producing the Oʁ#C sequences. Speakers almost never (≈0%) delete /ʁ/ and insert schwa at the same time. Only 20% of the tokens are aligned using the canonical pronunciation ([Oʁ#]), which correspond to 3-consonant sequences violating SSP ([OʁC]). The examination of these cases shows interesting results related to the type of the following consonant: more than 30% of the [Oʁ] pronunciation is followed by [l] (eg. quatre livre [katʁ#livʁ]). As illustrated in Figure 2, the distribution of the variants depends on speech style ($p< 0.001$). /ʁ/ deletion ([kat]) tends to be much more frequent than schwa insertion ([katʁə]) in the casual speech corpus NCCFr, whereas the opposite trend is found for the formal journalistic speech corpus ESTER. The less formal the speech style, the less we observe schwa insertion and the more we observe /ʁ/ deletion.

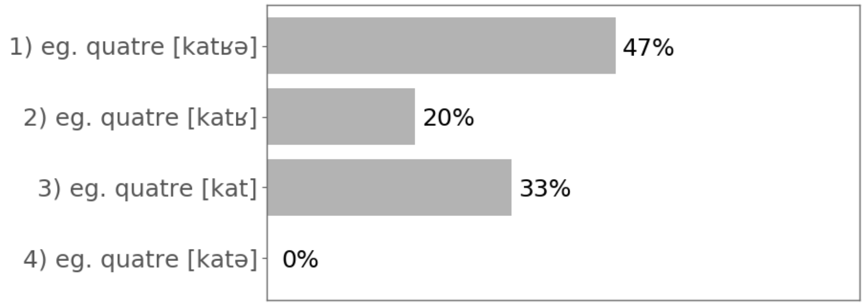


Figure 1. Distribution of the variants of the /Oʁ#/ cluster when followed by /#C/ over all corpora 1) eg. quatre [katʁə], 2) quatre [katʁ], 3) quatre [kat] 4) quatre [katə] indicate whether the word is realized 1) with epenthetic schwa, 2) canonically, 3) without word-final /ʁ/ and 4) without word-final /ʁ/ and with epenthetic schwa.

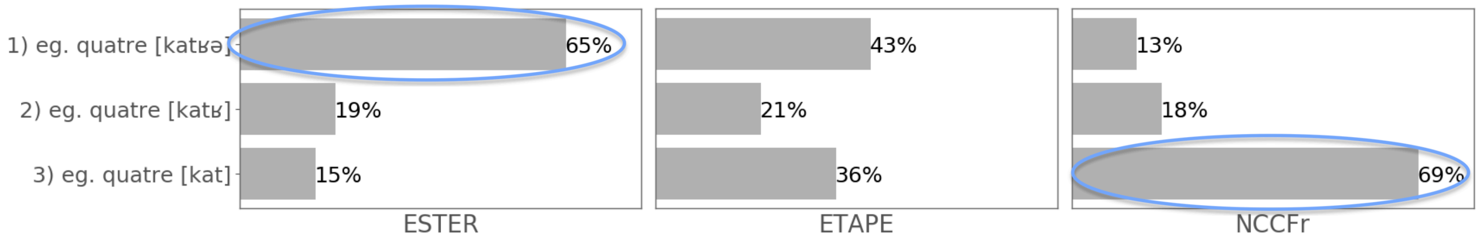


Figure 2. Distribution of the variants of /Oʁ#/ cluster when followed by /#C/ for each speech style.

[1] Clements, G. N. (1988). The sonority cycle and syllable organization. *Phonologica*, 63-76.

[2] Clements, G. N. (1990). The role of the sonority cycle in core syllabification. Papers in *laboratory phonology*, 1, 283-333.

[3] Côté, M. H. (2004). Consonant cluster simplification in Québec French. *Probus*, 16(2), 151-201.

[3] Laks, B. (1977). Contribution empirique à l'analyse socio-différentielle de la chute de /r/ dans les groupes consonantiques finals. *Langue française*, (34), 109-125.

[4] Galliano, S., Geoffrois, E., Gravier, G., Bonastre, J. F., Mostefa, D., & Choukri, K. (2006). Corpus description of the ESTER evaluation campaign for the rich transcription of French broadcast news. In *Proceedings of LREC* (Vol. 6, pp. 315-320).

[5] Gravier, G., Adda, G., Paulson, N., Carré, M., Giraudel, A., & Galibert, O. (2012). The ETAPE corpus for the evaluation of speech-based TV content processing in the French language. In *LREC-Eighth international conference on Language Resources and Evaluation*.

[6] Torreira, F., Adda-Decker, M., & Ernestus, M. (2010). The Nijmegen corpus of casual French. Speech Communication, 52(3), 201-212.

[7] Gauvain, J. L., Lamel, L., & Adda, G. (2002). The LIMSI broadcast news transcription system. *Speech communication*, 37(1), 89-108.

[8] Adda-Decker M., Lamel L. (2000). The use of lexica in automatic speech recognition. In *Lexicon Development for Speech and Language Processing*, F. van Eynde & D. Gibbon (Eds.), Kluwer Academic Publishers.

[9] Gamer, M., Lemon, J., Fellows, I., & Singh, P. (2012). irr: Various coefficients of interrater reliability and agreement. R package version 0.84, 137.

[10] New, B., Brysbaert, M., Veronis, J., & Pallier, C. (2007). The use of film subtitles to estimate word frequencies. *Applied psycholinguistics*, 28(4), 661-677.

[11] Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. arXiv preprint arXiv:1406.5823.

Variation of sibilants in three Ladin dialects

Yifan Yang[1], Rachel Walker[1] and Alessandro Vietti[2]
[1]University of Southern California, [2]Free University of Bozen-Bolzano

Overview. This paper presents an acoustic study of Ladin, a threatened minority Romance language spoken in Northeastern Italy, with a focus on the sibilants of three dialects (Brach, Cazet, and Moenat) spoken in the Fassa Valley of Trentino. The contributions are threefold. First, it provides up-to-date phonetic data for younger-generation speakers. Second, it reveals the nature of phonetic variation across dialects. Third, Smoothing Spline ANOVA[1] is used in the analysis of fricative spectral envelopes, an innovative application of this method, which provides insight on the nature of the differences amongst sibilants within and across dialects. It thereby exhibits a benefit of using statistical methods in the study of threatened or endangered languages.

Data collection. Previous research has identified two series of voiced and voiceless sibilant fricatives in Fassa dialects (denti-)alveolar and postalveolar. However, it does not show consensus on the nature of the post-alveolar series, which have been variously characterized as palatal(ized) or retroflex[2], [3], [4]. To investigate the nature of these sibilants acoustic recordings of Ladin words, produced in controlled conditions, were acquired in June 2018 in Vigo di Fassa. Four speakers spanning three dialects of Fassa Ladin were recorded: Two speakers of Brach (both F), one speaker of Cazet (F), and one speaker of Moenat (M). The speakers fell within an age range of 18-35. For the purpose of the current study, only words containing voiceless sibilants in our database were used for acoustic analysis, exemplified in (1). We note that these analyses serve as a preliminary study, since the sibilants used for spectral analysis were not controlled for syllable position and neighboring vowel quality.

Analysis and Results. The statistical method Smoothing Spline ANOVA (SSANOVA) has been utilized in phonetic studies for fitting ultrasound trajectories,[5] vowel formants,[6] liquid formants,[7] and fundamental frequency,[8] while this study uses the technique to provide a comparative illustration of the acoustic properties of the sibilants. For each token, a 10-ms window in the middle of the sibilant was selected, and the spectral envelope of the window was extracted. For each dialect, a SSANOVA model was fitted to the extracted spectral envelopes of the sibilants, shown in Figure 1; the interaction plots are shown in Figure 2. Since the post-alveolar series in Moenat was documented as retroflex, the comparison of the post-alveolar in three dialects is given in Figure 3.

Discussion. Within each dialect, the spectral envelope of the alveolar sibilant is significantly different from that of the post-alveolar since only a small portion of overlap can be observed between 5,000 Hz and 10,000 Hz in Figure 1. These results suggest that the two sounds are distinctive. Across dialects, however, the alveolar sibilant in Cazet stands out due to the peak it presents at approximately 7000 Hz (Figure 1b), indicating that it is more retracted compared to its counterparts in Brach and Moenat, and which renders it more similar spectrally to the post-alveolar. The post-alveolar fricative in Cazet shows a plateau between around 3500 Hz and 7000 Hz (Figure 1b). We suggest the possibility that post-alveolar sibilants could have two allophones characterized by two peaks, one at 3500 Hz and the other at around 7000 Hz. The variations could be contextual, which might be related to the syllable position of the post-alveolar or its following vowel. For instance, in words such as *dasc* [daʃ] 'give 2.SG.PRS' versus *scial* [ʃial] 'shawl', the post-alveolar either appears in coda position or in onset position before *i*. These contexts might result in the spectral variations seen in Cazet post-alveolar sibilants. Further, for the post-alveolar series, the noise energy peak in Moenat has lower frequency compared to the other two dialects, shown in Figure 3, which might indicate the backness and retroflex nature of this sound.[9] The results of our study are suggestive that the three Fassa dialects under study have developed post-alveolar sibilants that are each distinct from one another. This research provides a basis for future in-depth investigation into the properties of sibilants in Ladin.

(1) Word list (only some examples are listed)

	Brach		Cazet		Moenat	
alveolar	6 words; 28 tokens		10 words; 20 tokens		21 words; 41 tokens	
	das	[das]	das	[das]	asenz	[aˈsɛns]
	ors	[ors]	ors	[ors]	filz	[fils]
	sauch	[saˈuk]	sauch	[saˈuk]	son	[son]
	soreie	[soˈrɛie]	soreie	[soˈrɛie]	sort	[sɔrt]
post-alveolar	6 words; 25 tokens		8 words; 16 tokens		39 words; 106 tokens	
	dasc	[daʃ]	dasc	[daʃ]	dasc	[daʂ]
	orsc	[orʃ]	orsc	[orʃ]	orsc	[orʂ]
	scial	[ʃal]	scial	[ʃal]	stolz	[ʂtols]
	scigol	[ˈʃigol]	scigol	[ˈʃigol]	scaji	[ˈʂkazi̯]

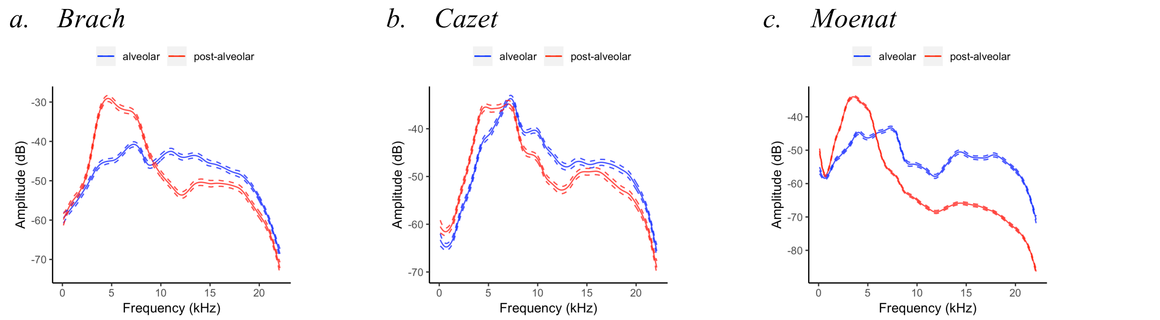


Figure 1. *Spectral information of sibilants in three dialects. The splines were plotted as the solid lines, while the dashed lines indicate the upper and lower boundary of 95% Bayesian confidence interval.*

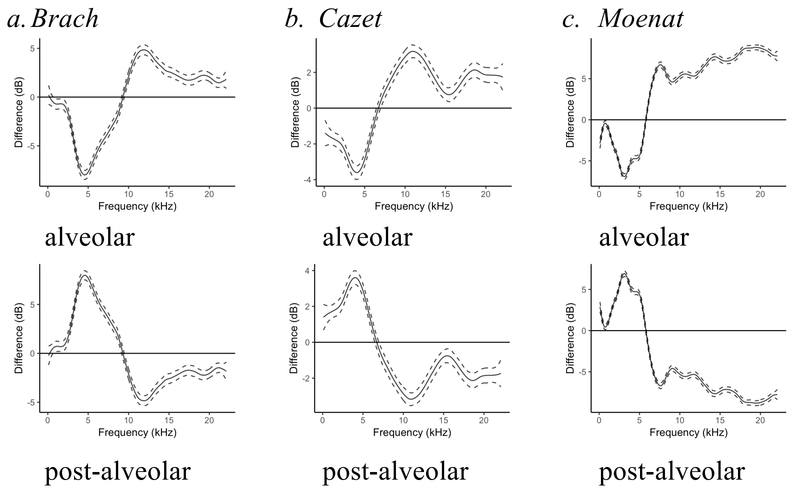


Figure 2. *Interaction effects with Bayesian confidence interval. The dashed lines indicate the upper and lower boundary of 95% Bayesian confidence interval.*

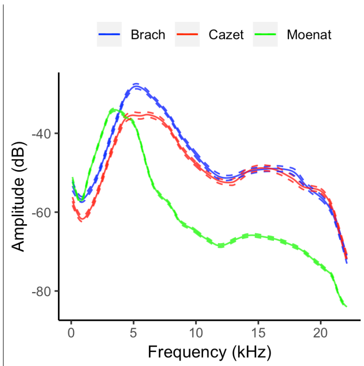


Figure 3. *Post-alveolar sibilant in three dialects. The post-alveolar in Moenat has a peak of lower frequency, indicating longer anterior cavity.*

[1] Gu, C. 2002. *Smoothing Spline ANOVA Models*. New York: Springer. **[2] Heilmann, L**. 1955. *La Parlata di Moena*. Bologna: Zanichelli. **[3] Chiocchetti, A**. 2017. Muamenti fonetici e fonematici nel ladino fassano dagli anni ’60 ad oggi. Mondo Ladino (41). 13-90. **[4] Salvi, G**. 2016. Ladin. In The Oxford Guide to the Romance languages, ed.by A. Ledgeway & M. Maiden, pp. 154-168. Oxford University. **[5] Davidson, L.** 2006. Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *JASA* 120(1). 407–415. **[6] De Decker, P. & J. Nycz**. 2006. A new way of analyzing vowels: comparing formant contours using smoothing spline ANOVA. Poster presented at NWAV 35, Columbus, OH. **[7] Howson, P**. 2018. Palatalization and rhotics: an acoustic examination of Sorbian. *Phonetica* 75. 132–150. **[8] Mathes, T. K. & A. Monthusi Chebanne**. 2018. An SS ANOVA (Smoothing Spline Analysis of Variance) study of high tone lowering in Tsua. *South African Journal of African Languages* 38(2). 137–148. **[9]. Gordon, M, P. Barthmaier & K. Sands**. 2002. A cross-linguistic acoustic study of voiceless fricatives. *JIPA* 32(2). 141–174.

Oral session 6

Factors affecting perception II

Perceptual evaluation of post-focal prominence in Italian by L1 and L2 naïve listeners

Caterina Ventura[1], Martine Grice[1], Michelina Savino[2], Aviad Albert[1] and Petra B. Schumacher[1]
[1]*University of Cologne,* [2]*University of Bari*

It is claimed that Italian is more rigid in the position of focus in comparison to West Germanic languages (see the continuum proposed by Face and D'Imperio [1]), with a larger tendency for the former to change the word order to maintain prominence in sentence-final position. Moreover, in this position Italian reportedly resists deaccenting words that are out of focus [2, 3], and it provides cues for post-lexical prominence in the post-focal domain even when this is not associated with any pitch movement [4].

The study explores the perception of prosodic prominence in Italian utterances as judged by Italian listeners, and by German listeners learning Italian. The aim is to ascertain whether the patterns of prominence perceived by native Italian listeners differ from those perceived by German listeners, given that the attenuation of post-focal material in Italian is claimed to be realized in a less categorical way than in German. The stimuli presented to the judges, structured as in (1), were produced by a Bari Italian speaker in a previous reading task.

(1) Bisogna pesare$_{\text{verb}}$ la farina$_{\textbf{TARGET}}$ con la bilancia$_{\text{noun}}$ *(You need to weigh the flour with the scales)*

For each sentence, three different renditions were produced: (i) with target word in Broad Focus (BF), (ii) contrastive Narrow Focus (NF) on the target, or (iii) Post-Focal (PF, verb focussed; target part of the background). Subjects (16 native speakers of Bari Italian;18 German learners of Italian, intermediate level) listened to 30 stimuli (10 items × 3 conditions) and rated the perceived prominence of each of the three content words (henceforth "critical words") in the utterance, corresponding to the verb preceding the target, the target word and the noun following the target. Ratings were given by using a slider, adapting the Rapid Prosody Transcription method [5] to include continuous responses and to relate the answers for the target to the other critical words, to address the relational property of prominence.

This paper focusses on the perceptual evaluation of BF and PF utterances, as they are the most relevant when comparing Italian and German in relation to post-focal prominence. Interestingly, our results show that Italian listeners do not perceive any difference between BF and PF in the degree of prominence of the target word (Fig. 1, left panel), whereas Germans do, as they assign a higher degree of prominence to the target word in BF than PF (Fig. 1, right panel). The latter outcome is reflected in the observed acoustic cues to prominence: in all stimuli, the target word in BF had a falling F0 contour (H+L*), whereas in PF the contour was flat. We also measured the Periodic Energy Mass (PEM) [6] for the stressed syllable of the three critical words, where PEM corresponds to the sum integral of periodic energy (directly linked to the acoustic intensity of f_0 and thus indicating the strength of perceived pitch) and duration, reflecting the prosodic strength of the corresponding syllable. Relative PEM is calculated by assigning a value of 100 to the strongest syllable of the utterance and calculating the value of the other syllables in relation to it. As shown in Fig. 2, in PF the verb is the only critical word that features a pitch movement, and exhibits a distinctively high PEM distribution, which is not the case for the target word (or following noun). Therefore, it appears that the main acoustic cues to prominence for the target in PF are neutralized, whereas the target in BF exhibits a pitch movement and a corresponding high PEM. These acoustic observations support the interpretation of our prominence perception results as implying language-specific differences that are projected onto the L2: Germans appear to follow a more pitch-driven strategy, interpreting the flat pitch of post-focal targets as completely attenuated, as it would be in their native language [7]. By contrast, when estimating post-focal prominences Italian listeners seem to pay attention to factors other than pitch and energy, relying more on their native language-driven expectations to find cues for prominence in post-focal position.

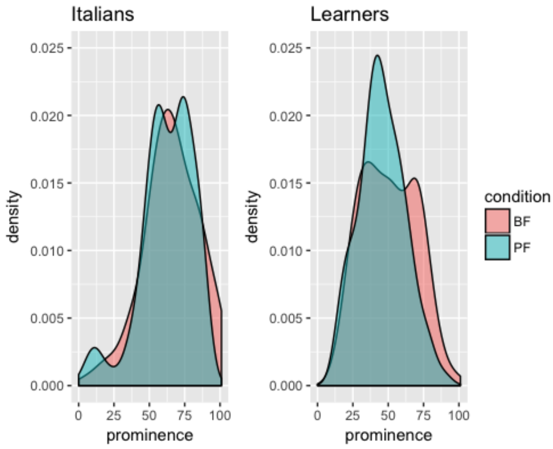


Figure 1. Prominence judgements of target word in Broad Focus (BF) and Post-Focal (PF) conditions by native Italians and German learners.

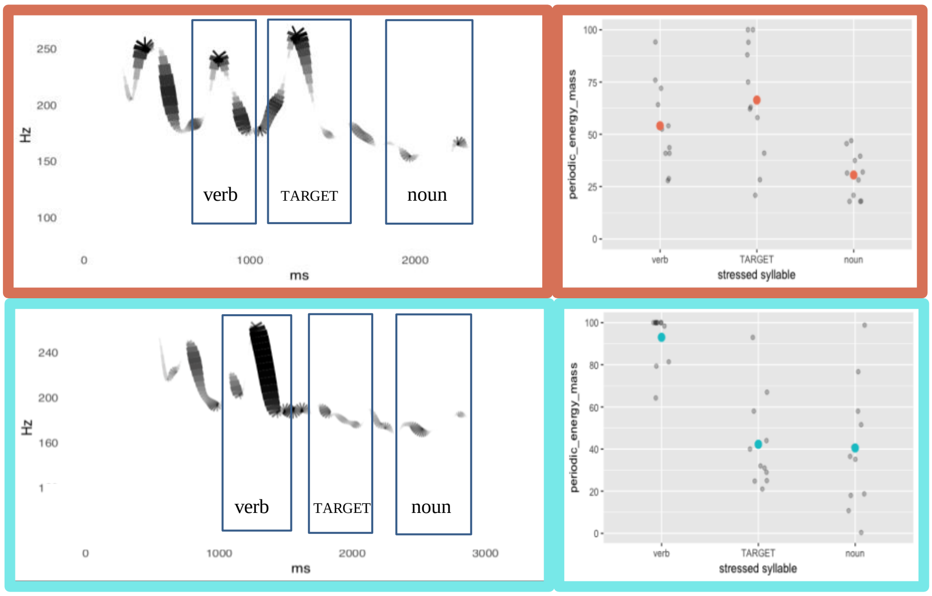


Figure 2. Left panels: Estimated perceived pitch contour (*Periograms* [6]) of the utterance "Bisogna pesare$_{verb}$ la farina$_{\textbf{TARGET}}$ con la bilancia$_{noun}$" in BF (upper panel) and PF (bottom panel). Right panels: Periodic Energy Mass for critical words in BF (upper panel) and PF (bottom panel).

[1] Face, T. L. & D'Imperio, M. 2005. Reconsidering a Focal Typology: Evidence from Spanish and Italian. In Pacini (Ed.), *Italian Journal of Linguistics / Rivista di linguistica*, 271-289.

[2] Grice, M., D'Imperio, M., Savino, M., Avesani, C. 2005. Strategies for intonation labelling across varieties of Italian. In Sun-Ah Jun (Ed.), *Prosodic Typology: the Phonology of Intonation and Phrasing*, New York: Oxford University Press, 362-389.

[3] 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.

[4] Bocci, G. & Avesani, C. 2011. Phrasal prominences do not need pitch movements: postfocal phrasal heads in Italian. *Proceedings of INTERSPEECH 2011*, Florence, 1357-1360.

[5] Cole, J., Mo, Y., & Hasegawa-Johnson, M. 2010. Signal-based and expectation-based factors in the perception of prosodic prominence. *Laboratory Phonology*, *1*, 425–452.

[6] Albert, A., Cangemi, F. & Grice, M. 2018. Using periodic energy to enrich acoustic representations of pitch in speech: A demonstration. *Proceedings of the 9th International Conference on Speech Prosody*, Poznan, 804-808.

[7] Röhr C. & Baumann, S. 2010. Prosodic marking of information status in German. *Proceedings of the 5th International Conference on Speech Prosody*, Chicago, USA, 1-4.

Multimodality in prominence production and its sensitivity for lexical prosody

Gilbert Ambrazaitis[1] and David House[2]
[1]*Linnaeus University, Växjö,* [2]*KTH (Royal Institute of Technology), Stockholm, Sweden*

Prominence is essentially a multimodal phenomenon, as verbal prominence markers such as pitch accents are frequently accompanied by so-called beat gestures, typically produced by the hands, the head, or the eyebrows [1-3] and visual beats seem to add to perceived prominence [4-6]. However, our understanding of gesture-speech integration is still far from complete. One important issue currently under discussion is whether the verbal and the visual modality tend to interact in a compensatory fashion in prominence production, where less verbal effort is required when a visual beat is added or vice versa, or whether multimodal prominence is rather cumulative in nature, with pitch accents and beat gestures reinforcing each other [3-7]. Another unresolved question is how visual beats interact with lexical-prosodic structure. So far, previous studies have only more or less implicitly suggested a link between beats and lexical prosody, as gesture strokes have been shown to associate and temporally align with lexically stressed syllables [2][8-9].

The present study addresses these two issues by examining the phonetic realization of pitch accents in Swedish as a function of accompanying beat gestures by the head and the eyebrows, as well as of the lexical-prosodic structure of the accented words. Concerning the latter, Swedish exhibits lexical stress and tone, which are both connected to the language's well-known "word accents" (Accent 1, Accent 2; henceforth, A1, A2): Tone is lexically specified only in A2 single-stress words, while it is post-lexical in A1 as well as in A2 words with additional secondary stresse(s) (compounds and some derivations) [10]. We can thus distinguish between three lexical-prosodic categories: (i) simplex stress (A1), (ii) simplex stress + tone (A2) and (iii) compound stress (A2). In addition, Swedish distinguishes between two tonal prominence levels, as we can distinguish between a "small" (HL – realized HL* for A1 and H*L for A2) and a "big" accent (HLH – realized (H)L*H or H*LH) (terminology adopted from [10]).

Our study is based on 12 minutes of audio and video data (1936 words) from five Swedish television news anchors (two female). The material was annotated for big accents (BA), head beats (HB) and eyebrow beats (EB) (Fleiss' kappa: $\kappa_{BA} = 0.77$; $\kappa_{HB} = 0.69$; $\kappa_{EB} = 0.72$), as well as for tonal targets. The realization of the falling (HL) and the rising part (LH) of big accents (fall/rise range in semitones) was examined in three lexical-prosodic conditions (see i-iii above), comparing three (multimodal) constellations of prominence markers: (a) words produced with a BA only (without a beat gesture), (b) words with BA co-occurring with a HB (BA+HB) and (c) words with BA and both HB and EB (BA+HB+EB).

The results suggest a slight tendency for a positive correlation (i.e. a cumulative relation) between the presence of visual beats and the realization of the BA-rise (Fig. 1a). For the accentual fall, the situation is less clear and more complex (Fig. 1b). For both measures, the results suggest a strong interaction between lexical prosody and the constellation of prominence markers. For instance, the combination of a head and an eyebrow beat, but not a head beat alone, seems to induce an enlarged BA-rise in simplex stress words (either A1 or A2), while for the compound stress words, an enlarged pitch range is observed even with the addition of a head beat alone (Fig. 1a). Crucially, the interaction between the multimodal prominence constellation and the lexical-prosodic condition is significant both for the BA-rise *($p=.017$)*[i], and the accentual fall *($p<.001$)*, while there is no significant main effect of the multimodal prominence constellation on either of the two measures. The results are in line with the idea of an essentially cumulative relation between verbal and visual prominence markers in production, at the same time suggesting an interaction of the two modalities that is sensitive for lexical prosody, an issue that needs to be further examined in future research.

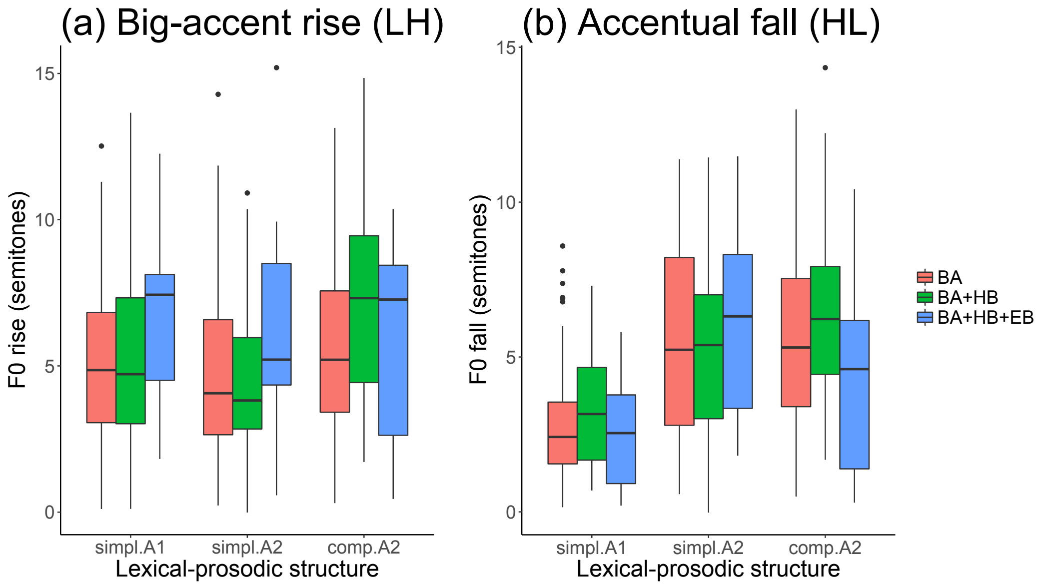


Figure 1. *Boxplots for the BA-rise (a) and the preceding accentual fall (b) measured in semitones as a function of the multimodal prominence constellation (colors) and the lexical-prosodic condition (x-axis: simplex stress (Accent 1), simplex stress + tone (Accent 2), compound stress (Accent 2)); n_{BA}=276, n_{BA+HB} =178, $n_{BA+HB+EB}$=73.*

[1] 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.

[2] Loehr, D. 2012. Temporal, structural, and pragmatic synchrony between intonation and gesture. *Journal of the Association for Laboratory Phonology* 3, 71-89.

[3] Ambrazaitis, G., & House, D. 2017. Multimodal prominences: Exploring the patterning and usage of focal pitch accents, head beats and eyebrow beats in Swedish television news readings. *Speech Communication* 95, 100-113.

[4] Krahmer, E., & Swerts, M. 2007. The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language* 57, 396–414.

[5] Dohen, M., & Loevenbruck, H. 2009. Interaction of audition and vision for the perception of prosodic contrastive focus. *Language and Speech* 52, 177–206.

[6] 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.

[7] Roustan, B., & Dohen, M. 2010. Co-Production of Contrastive Prosodic Focus and Manual Gestures: Temporal Coordination and Effects on the Acoustic and Articulatory Correlates of Focus. *Proceedings of Speech Prosody 2010* (Chicago, IL, USA).

[8] Leonard, T., & Cummins, F. 2011. The temporal relation between beat gestures and speech. *Language and Cognitive Processes* 26, 1457−1471.

[9] Esteve-Gibert, N., & Prieto, P. 2013. Prosodic structure shapes the temporal realization of intonation and manual gesture movements. *Journal of Speech, Language, and Hearing Research* 56 (3), 850-864.

[10] Myrberg, S., & Riad, T. 2015. The prosodic hierarchy of Swedish. *Nordic Journal of Linguistics* 23(2), 115-47.

[i] Tested by means of likelihood-ratio tests based on linear mixed regression models with the multimodal prominence constellation (3 levels), lexical-prosodic structure (3 levels), and speaker sex (2 levels) as fixed factors, assuming random intercepts and slopes for speaker.

Poster session 4

The effect of pitch accent on V-to-V coarticulation induced variability of vowels

Andrea Deme[1,2], Márton Bartók[1,2], Tekla Etelka Gráczi[3,2], Tamás Gábor Csapó[4], Alexandra Markó[1,2]

[1]Eötvös Loránd University, [2]MTA-ELTE „Lendület" Lingual Articulation Research Group,[3]Research Institute for Linguistics HAS, [4]Budapest University of Technology and Economics

Coarticulation is one of the main sources of segmental variability. Since the seminal work of [1] it is recognized that not only adjacent speech sounds but also transconsonantal vowels have an effect on each other, and the vowels in V_1CV_2 sequences are claimed to be produced with one single underlying diphthongal gesture to which the consonant's gesture is superimposed. The extent a segment is susceptible to coarticulation, i.e., the contextual variability it exhibits, is referred to as coarticulatory resistance (CR; greater resistance = less variance) [2]. CR in V-to-V coarticulation may be influenced by several factors. In an acoustic study [3] showed that Vs show smaller variability, if they are in a (lexically) **stress**ed syllable (vs. unstressed) (5 speakers). [4] confirmed that the above effect also exists for higher level (sentential) **accent** taking the edge and the first quarter of the V as points of measurement, but he tested it only in the articulatory domain (6 speakers). Although inconclusively and in smaller samples, it was also shown that certain **V-qualities** show greater resistance than others: in German, /i/ was found to be more resistant than /a/ (3 speakers) [5], in Italian, /i/ was more resistant than /a/, and /a/ than /u/ (1 speaker) [6], while in Thai, the high Vs /i/ and /u/ were similarly resistant (6 speakers) [7]. Lastly, [8] demonstrated that **intervening Cs** which exert a smaller degree of tongue dorsum contact with the palate allow for more V-to-V coarticulation (5 speakers). In an attempt (1) to further explore if prominence provokes CR in V-to-V coarticulation, (2) to uncover the language-specificity of the effect of prominence, and (3) to clarify the effect of V-quality, in the present study we analysed V-to-V carryover coarticulatory effects in the acoustic domain, in real words, in minimally constrained C-context (to maximize V-to-V effects), in the presence/absence of sentence level accent (+ word stress co-varying with accent) in Hungarian, and in a larger dataset (i.e., in more speakers) than previous studies.

We recorded 10 Hungarian adult female speakers producing /uh**u**/C_{alv}/u/, /ih**u**/C_{alv}/u/, /ih**i**/C_{alv}/i/, and /uh**i**/C_{alv}/i/ in words embedded in meaningful sentences, in two accent conditions: ˈV*h***V**C_{alv}V and V#ˈ*h***V**C_{alv}V. We used the glottal fricative /h/, as it is underspecified for oral configuration, and thus interferes the least with the single diphtongual gesture of the V segments. We measured F_1 and F_2 of V_2 at the left edge (median of first 10%; F_{2onset}) and in the temporal midpoint (median of mid 10%; F_{2mid}). Building on the locus equation approach, to gauge the **degree of coarticulation,** we fitted linear models on F_{2mid} and F_{2onset}, as a function of the tested variables [2]. **V-variability** was quantified by i) the magnitude of V target dispersion expressed in Euclidean distances of V_2 data points from the midpoint of the V ellipses in the $F_{1mid} \times F_{2mid}$ plane (separately for /i/ and /u/ × accent cond. × context), and ii) the difference of F_{2onset}s of coarticulated and non-coarticulated instances. The latter two measures were tested with linear mixed effects models.

Steep slopes for /i/ and slopes of approx. 0 for /u/ in both conditions reflect that /i/s were produced more stationary in time than /u/s, irrespective of the presence of accent (Fig 1). The analysis of Euclidean distances revealed that tokens were more variable in /i/ than in /u/ [$F(1, 10) = 10.55$, $p < 0.01$)], and in symmetrical (vs. asymmetrical) contexts [$F(1, 14) = 10.33$, $p < 0.01$)] irrespective of the presence of accent (Fig 2). This finding along with the regression fits suggests that the more dynamic realization of /u/ tokens resulted in higher accuracy in reaching V-target in /u/. Finally, F_{2onset} differences showed that there is generally little difference between coarticulated and non-coarticulated Vs, but in unaccented condition, /i/ varied more due to coarticulation (V × condition interaction: [$F(1, 30) = 16.04$, $p < 0.01$)]).

These results contradict some of the previous findings on V-quality, as we found that /u/ showed less variation than /i/. Moreover, results partly also contradict [3, 4] with respect to the effect of pitch accent, as we found that the lack of accent decreased CR only in /i/. The striking divergence of results may stem from the numerous methodological differences of the cited studies and the present paper (i.e., maximised V-to-V effect, use of real words, different quantification of variance), and the larger sample size used in this study, but may also point to language specific patterns in the interaction of prosody and V-to-V induced variation.

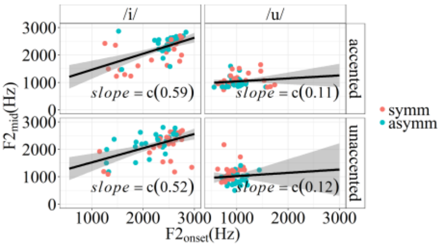


Figure 1. "Locus equations" for the target V_2 in coarticulating (asymm) and non-coarticulating (symm) contexts, as a function of prominence

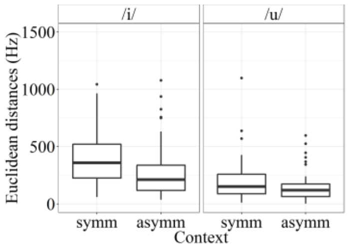


Figure 2. Acoustic dispersion of /i/ and /u/ on the basis of V_{2mid} spectral values

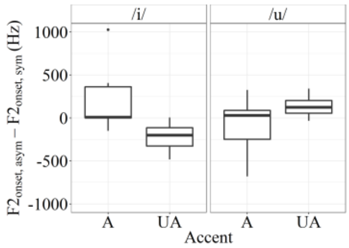


Figure 3. Differences of F_{2onset}s of coarticulated (asym) and non-coarticulated (symm) vowels

[1] Öhman, S. E. G. 1966. Coarticulation in VCV utterances: spectrographic measurements. *Journal of the Acoustical Society of America* 39, 151–168.

[2] Bang, H.-Y. 2017. The acoustic counterpart to coarticulation resistance and aggressiveness in locus equation metrics and vowel dispersion. *The Journal of the Acoustical Society of America* 141. EL345–EL350.

[3] Fowler, C. A. 1981. Production and perception of coarticulation among stressed and unstressed vowels. *Journal of Speech and Hearing Research* 46. 127–139.

[4] Farnetani E.,Vagges K. & Magno-Caldognetto E. 1985. Coarticulation in Italian /VtV/ Sequences: A Palatographic Study. *Phonetica* 42. 78–99.

[5] Cho, T. 2004. Prosodically conditioned strengthening and vowel-to-vowel coarticulation in English. *Journal of Phonetics* 32. 141–176.

[6] Butcher, A. & Weiher, E. 1976 An electropalatographic investigation of coarticulation in VCV sequences. *Journal of Phonetics* 4. 59–74.

[7] Farnetani, E. & Recasens, D. 2010. Coarticulation and connected speech processes. In Hardcastle, W. J., Laver, J. & Gibbon, F. E. (eds.) *Handbook of Phonetic Sciences. (2nd Ed.)* Chichester, U.K.: Wiley-Blackwell.

[8] Mok P. P.K. 2011. Effects of vowel duration and vowel quality on vowel-to-vowel coarticulation. *Language and Speech* 54. 527–45.

[9] Recasens, D. 2015. The effect of stress and speech rate on vowel coarticulation in Catalan vowel–consonant–vowel sequences. *Journal of Speech, Language, and Hearing Research* 58. 1407–1424.

While conducting the study, the first author was supported by the *Hungarian Ministry of Human Capacities (Emberi Erőforrások Minisztériuma), Hungarian Talent Program (Nemzeti Tehetség Program) NTP-NFTÖ-18.*

Tune choice in utterance-initial vocatives in Asturian

Eduardo García-Fernández
University of Massachusetts Amherst

Vocative marking in Asturian, a minority Ibero-Romance language spoken in Asturias (North Western Spain), is realized both intonationally and morphosyntactically in utterance-initial position. Intonationally vocatives are realized with one of two intonation contours: L+H* (as in Figure 1) or H+L* (Figures 2 & 3). Morphosyntactically they can be marked by an optional phrase-initial particle *a-*, for example *¡A Manolo!* (Manolo!). A previous production experiment confirmed that the implementation of the H+L* pitch accent is subject to phonological restrictions that affect the presence or absence of the particle. There is evidence that monosyllabic nouns and words with word-initial stress do not provide the necessary pre-tonic material that the leading H tone requires to align with (as in 2a below). In such cases, H+L* implementation is licensed by the addition of the particle a- to the vocative noun, which automatically provides the leading tone (H) with an anchoring location, as in 2b (text-to-tune accommodation). However, the particle a- can precede any vocative noun, regardless of stress placement (example 2d) and is intrinsically associated with H+L* (therefore incompatible with L+H*). In a previous experiment, the author found that, in an imitation task, participants failed to (re)produce vocative calls represented in examples 1b, 1d, and 2a.

1. [L+H*]:
a. *Mari, …*
b. *#A Mari, …*
c. *Manolo, …*
d. *#A Manolo, …*

2. [H+L*]:
a. *#Mari, …*
b. *A Mari, …*
c. *Manolo, …*
d. *A Manolo, …*

The purpose of this study was to uncover the pragmatic restrictions that guide Asturian speakers in tune choice. Cross-linguistically, various melodies may occur for the same utterance type and, oftentimes, this variation can be explained by the different types of epistemic stances that speakers can take. Intonation can convey relations between propositional content of utterances and the mutual belief space between speaker (S) and hearer (H) (Pierrehumbert & Hirschberg 1990). Recently, the relation between intonation, propositional content (*p*), and beliefs has been investigated in polar questions (e.g., Vanrell et al. 2014; Armstrong & Prieto 2015), declaratives (e.g., Gravano et al. 2008), and imperatives (Armstrong & Lesho 2016). In these sentence types, interlocutors naturally navigate the mutual belief space, but could Asturian vocatives show similar behavior? In an open-response task, the intuitions of 26 native speakers revealed that H+L* is used in contexts where there is a mismatch in the interlocutors' mutual belief space or common ground (CG). This gap can be either S-oriented (S's surprise/counter-expectation/doubt), or H-oriented (S is reproaching/criticizing/impatient/giving an order). On the contrary, L+H* does not convey those meanings. In order to investigate whether the use of H+L* is more likely to be interpreted as implicating that some *p* is outside of the interlocutors' CG, 15 native Asturian speakers participated in a perception task, where they listened to utterance-initial vocatives, produced either with L+H* or H+L*, and followed by 2 possible completing utterances: one pragmatically biased with a mismatch in CG (either S- or H-oriented), and the other unbiased. They were asked to choose the utterance that would follow each vocative more suitably. The results (Graph 1) reveal that the use of H+L* in utterance-initial vocatives in Asturian arises pragmatically in contexts where S marks a gap in the mutual belief space, either on S or H side. In a larger picture, this study contributes to what we know about intonational meaning. As it is the case with other utterance types (e.g., polar questions), vocatives can also reference belief states through intonation. In Asturian, utterance-initial vocatives are open for marking a gap in the interlocutors' CG.

Figures 1, 2, and 3. Waveform, spectrogram, and F0 track of Asturian vocative calls *¡Manolo!,…* (L+H* [left]; H+L* [center]) and *¡Á Monolo!,* … (H+L* [right])

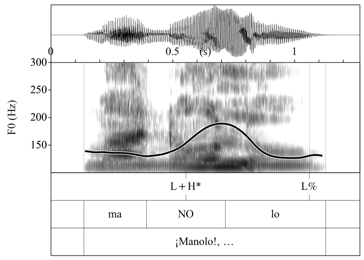


Figure 1

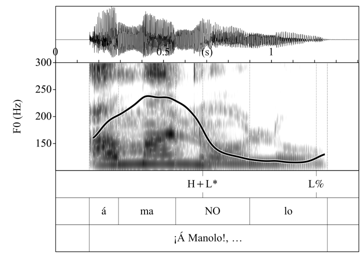


Figure 2

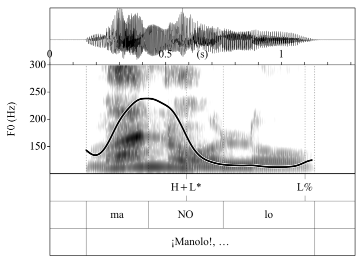
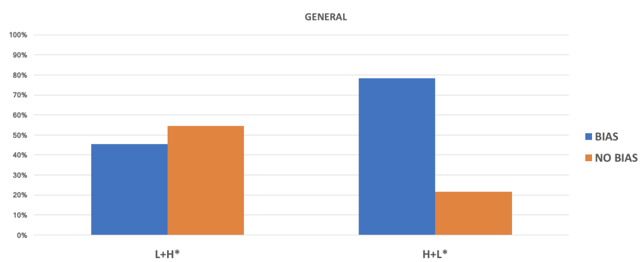


Figure 3

Graph 1. Results from the perception task showing the percentage of biased and unbiased responses associated with utterance-initial vocatives produced with L+H* and H+L* respectively.



Graph 1

References

Armstrong, M. E., & Prieto, P. (2015). The contribution of context and contour to perceived belief in polar questions. *Journal of Pragmatics*, 81: 77–92. doi: 10.1016/j.pragma.2015.03.006

Armstrong, M.E., & Lesho, M. (2016). Intonationally-encoded implicature and regional variation in American English imperatives. Poster presented at *Linguistic Society of America Annual Meeting 2016*. Washington, D.C.

Gravano, A., Benus, S., Hirschberg, J., German, E. S., & Ward, G. (2008). The effect of contour type and epistemic modality on the assessment of speaker certainty. *Proceedings of Speech Prosody* – Campinas: 401-404.

Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. Cohen, J. Morgan, & M. Pollack(Eds.), *Intentions in communication* (pp. 271–311). Cambridge MA: The MIT Press.

Vanrell, M. M., Armstrong, M.E., & Prieto, P. (2014). The role of prosody in the encoding of evidentiality in Catalan. *Proceedings of Speech Prosody 2014*:1022-1025.

On the prosodic cues of ironic exclamatives. A pilot study.

Glenda Gurrado
University of Bari, Italy

Irony is traditionally considered a semantic inversion between the primary meaning and the implicated one: it is defined as a matter of saying one thing and meaning something else. Nevertheless, to date a number of alternative interpretations of irony have been proposed, like the echoic account and the pretence theories (for additional references, see also [1]). Irony has different aims but, basically, it is an attitude strictly connected to the willing of safeguarding the personal image: by means of irony speakers can hide their real intention blurring the lines between context and content. With reference to its purposes, two main categories of irony can be distinguished: sarcasm and teasing. Sarcasm is used to communicate an impolite valuation of the addresse using a normally polite sentence, on the contrary teasing indirectly communicates a positive valuation by means of a verbal criticism [2].

The role of context is essential for the production and the identification of an ironic utterance. The pragmatic dimension of irony is, however, strongly connected to the prosodic one. When contextual elements are ambiguous or unavailable, the prosodic cues of verbal irony become more perceptible [3]. To date, relatively few studies have focused on the acoustic analysis of ironic speech. Nasal articulation, reduced speech rate, heightened pitch variation, but also changes in voice quality and monotonic or lowered pitch (flat irony) are some of the typical prosodic cues of ironic speech identified so far (among others [4], [5], [6], [7]). Most recent studies have confirmed these findings and highlighted other cues that strongly mark sarcastic voice [3], [8]. However, with specific reference to Italian, there is currently little available research focused on the acoustic description of verbal irony [2], [9], [10].

On the basis of these issues, the main goal of this research is to investigate the differences between literal and sarcastic utterances in Bari Italian in terms of acoustic parameters. To pursue this aim, we selected exclamative sentences: we opted for this phrasal type because in spoken Italian sarcastic attitude is usually communicated by means of exclamatives.

For our research, six male speakers of Bari Italian (24-33 years old) were invited to read six pairs of texts. The two texts of each pair described two different situations both ending with the same exclamative sentence: the context defined the attitude (sarcastic or sincere) that each speaker would have used to utter the sentence. An acoustic analysis of the speech material was carried out, considering the variables employed in this field of study: average $F0$x, minimum $F0$, maximum $F0$, $F0$ range, onset and offset $F0$ values, duration, speech rate, mean intensity and the duration of the last stressed vowel. The data collected were subjected to statistical analysis. The results showed that some parameters are more involved in the characterization of verbal irony than others: higher minimum, maximum and offset $F0$ values, a higher intensity, a slower speech rate and a longer duration of the last stressed vowel seemed to characterize sarcastic exclamatives. Nevertheless, results also confirmed that different vocal profiles are related to the communication of sarcasm. In the present research we identified two main tendencies: three of the speakers realized flat irony, while the other three showed a more heightened pitch variation.

The present study highlights the extreme variability of irony, an attitude that allows to play with voice in order to indirectly communicate a positive or a negative intention.

[1] Wilson, D. 2017. Irony, Hyperbole, Jokes and Banter. In Blochiwwiak, J., Grisot, C., Durrieman, S. & Laenzliger, C. (Eds.), *Formal Models in the study of languages : Application in Interdisciplinary Context*. Cham: Springer, 201-220.

[2] Anolli, L., Infantino, M., & Ciceri, R. 2002. From "Blame by Praise" to "Praise by Blame": Analysis of vocal patterns in ironic communication. *International Journal of Psychology* 37 (5), 266-276.

[3] Mauchand, M., Vergis, N., & Pell, M. 2018. Ironic tones of voices. *Proceedings of the 9th International Conference on Speech Prosody 2018* (Poznań, Poland), 443-447.

[4] Fonagy, I. 1971. Synthèse de l'ironie. *Phonetica* 23, 42-51.

[5] Cutler, A. 1974. On saying what you mean without meaning what you say. In LaGaly, M.W., Fox, R.A., Bruck, A. (Eds.), *Papers from the tenth regional meeting of Chicago Linguistic Society*. Chicago: Chicago Linguistics Society, IL, 117-127.

[6] Mueke, D.C. 1978. Irony markers. *Poetics* 7, 363-375.

[7] Schaffer, R. 1982. Are there consistent vocal clues for irony? In Masek, C.S., Hendrick, R.A., & Miller, M.F. (Eds.), *Parasession on Language and Behavior*. Chicago: Chicago Linguistic Society, IL, 204-210.

[8] Cheang, H. S., & Pell, M. D. 2008. The sound of sarcasm. *Speech Communication* 50 (5), 366-381.

[9] S. Attardo, J. Eisterhold, Hay J. and I. Poggi, "Multimodal markers of irony and sarcasm," *Humor: International Journal of Humor Research* 16, n.2, pp. 243-260, 2003.

[10] Gili Fivela, B., & Bazzanella, C. 2014. The relevance of prosody and context to the interplay between intensity and politeness. An exploratory study on Italian. In Hidalgo Navarro, A. (Ed.), *Special Issue: The prosodic expression of linguistic im/politeness in Romance Languages. Journal of Politeness Research* 10, n.1, 97-126.

The prosody of impersonating characters in storytelling speech

Anna Huszár
Research Institute for Linguistics, Hungarian Academy of Sciences

Storytellers capture their audience's attention by varying their voice. They use prosody to impersonate the different characters. Previous studies indicate that storytellers use much more variation in pitch and intensity than other speakers, for example newsreaders [1, 2]. They also found that storytellers tend to speak slower and take longer pauses than newsreaders. Another study compared storytellers' speech to neutral speech [3] and observed that storytellers' average pitch and intensity values are larger than in neutral style. They also found that neutral style is faster than storytellers' speech. Doukhan et al. [4] showed that changes in pitch and intensity are relevant to impersonate different kinds of characters.

The phenomenon, however, has not been studied for Hungarian. The purpose of this study is to examine which prosodic parameters can play a significant role in impersonating different characters. According to the main hypothesis the participants change the articulation rate and the pitch to impersonate the characters.

Eight female students (17–18-year-olds) specialized in acting from a Budapest secondary grammar school were recorded. A text (consisting of 40 sentences) was written for the experiment. The participants were instructed to read the text five times: first in neutral style and then they were instructed to impersonate four different kinds of characters. The characters were the following: the excited hamster, the lazy sloth, the graceful deer and the sly fox. A drawing depicting the characters was given to the participants to help them to impersonate the characters. The annotation of the recorded text was carried out by using Praat [5]. The sentences, the pause-to-pause intervals and the irregular parts were segmented manually. The duration was extracted using a Praat script and articulation rate was calculated by sentences and measured as the number of syllables by seconds (pauses excluded). Pitch was also extracted using a Praat script (irregular parts were excluded), f0-mean and f0-range were analysed.

Preliminary results show that changes of pitch and articulation rate are relevant to impersonate the different characters. Faster articulation tempo was found in the case of the excited hamster. The other three characters were linked with slower articulation rate, the slowest tempo was measured in the case of the lazy sloth. The participants tend to exhibit a higher pitch when impersonating the excited hamster or the graceful deer and a lower pitch when impersonating the lazy sloth or the sly fox.

The study proved that Hungarians also use pitch and articulation rate to impersonate different characters. Future work will consist of investigating what other prosodic features are used for impersonating characters.

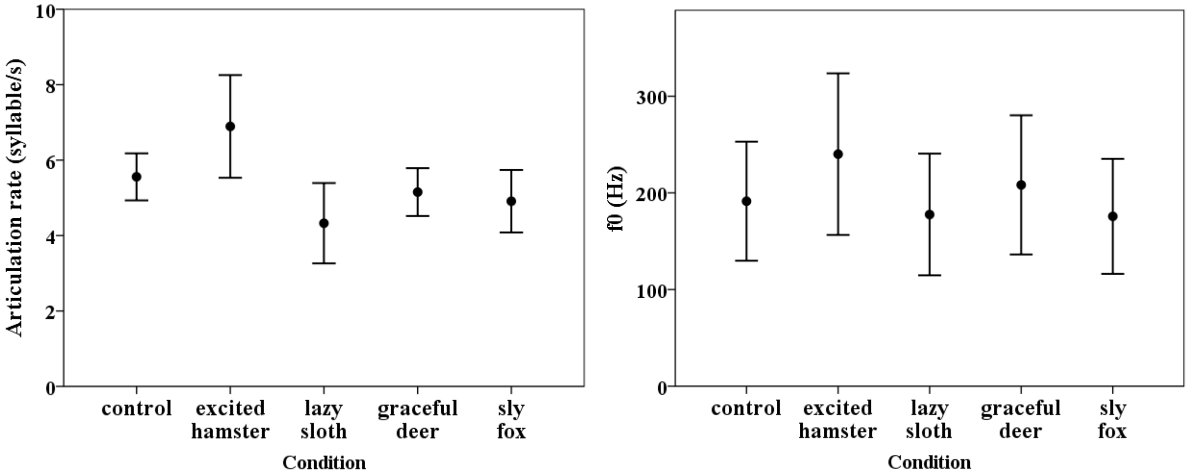


Figure 1. *The mean and the standard deviation of articulation rate and f0 across the conditions.*

[1] Fackrell, J., Vereecken, H., Buhmann, J., Martens J-P., & Van Coile, B. 2000. Prosodic variation with text type. *Proceedings of 6th International Conference on Spoken Language Processing (ICSLP 2000), Vol.3*. 231–234.

[2] Theune, M., Meijs, K., Heylen, D., & Ordelman, R. 2006. Generating expressive speech for storytelling applications. *Speech, and Language Processing, 14(4)*. 1137–1144.

[3] Verma, R., Sarkar, P., & Rao, K., 2015. Conversion of neutral speech to storytelling style speech. *Advances in Pattern Recognition (ICAPR)*. 1–6.

[4] Doukhan, D., Rilliard, A., Rosset, S., Adda-Decker, M., & d'Alessandro, C. 2011. Prosodic analysis of a corpus of tales. *Interspeech-2011*. 3129–3132.

[5] Boersma, P., & Weenink, D. 2014. *Praat: doing phonetics by computer*. http://www.fon.hum.uva.nl/praat/download_win.html

Cluster simplification in Russian children with Specific Language Impairment

Darya Kavitskaya, UC Berkeley
Sharon Inkelas, UC Berkeley

Much data on the behaviour of clusters in nonword repetition tasks [3] are drawn from Germanic languages (primarily English), with tightly restricted cluster possibilities. For instance, previous studies of cluster simplification in the productions of young children have found a tendency to preserve the obstruent in onset clusters and the sonorant in coda clusters; [2] and [4] suggest this may be due to universal syllable markedness. However, Russian is more permissive: unlike in English, Russian onset and coda clusters can both increase and decrease in sonority and thus, both unmarked and marked clusters are attested. This paper reports results from a nonword repetition study that addresses the strategies of cluster simplification in two matched Russian-speaking child populations aged 7-10, one typically developing (TD) and one with specific language impairment (SLI). Subjects were matched for age and all had IQs in the normal range. The findings support the view that the nature of phonological deficit in SLI is a phonological short-term memory impairment and consider implications of cluster simplification.

Method. The stimuli consisted of 144 disyllabic nonwords. Stimuli were designed to test repetition accuracy of consonant clusters of various sizes, varying sonority, and in different word positions (initial or final). 72 non-words had CC clusters; 36 had CCC clusters, and the remainder had singleton consonants. Clusters were divided evenly among onsets and codas, and among rising, level, and falling sonority (1). Clusters that obey the Sonority Sequencing Generalization (SSG), that is, onsets of rising sonority and codas of falling sonority were coded as marked, while the others were coded as unmarked. Half of CC clusters were attested in actual Russian words, and none of the CCC clusters used in the stimuli were attested.

Results. Overall, SLI children were less accurate than TD children (Chi-square test, $p < .008$). SLI subjects make more errors in clusters than TD subjects do ($p < .003$). Larger-sized CCC clusters were more prone to error in both groups than smaller-sized CC clusters (for TD, $p < .001$, for SLI, $p < .001$). While TD subjects were more accurate in attested clusters ($p < .01$), for SLI subjects, cluster attestedness was not significant.

Errors in the production of clusters ranged over deletion, epenthesis, metathesis, assimilation, dissimilation, substitution, and reduplication (2). This study focuses on deletion, which was the most common repair for both groups of children. The study showed that children with SLI delete more than TD children. For both groups, deletion of a consonant adjacent to vowel was inhibited.

Implications. This study has several important implications. First, as documented for a related population by [5], the differences in production patterns between TD and SLI children are essentially quantitative, not qualitative. Children with SLI exhibit a higher error rate overall, but the main tendencies are similar. Nonword repetition has been used in the past as a measure of phonological memory [3], [8]. Our results support the view that the phonological deficit in individuals with SLI involves decreased phonological short-term memory, not a restriction to the most unmarked (CV) syllable structure [6].

Second, it is important to consider languages with a diverse set of onset and coda clusters when formulating theories of cluster reduction. If syllable markedness determines the identity of the consonant surviving a cluster simplification process, then, unlike in the Russian data, consonant type, not position, should be the main predictor. Our interpretation of the tendency to accurately reproduce the vowel-adjacent member of a cluster is that children are most accurate at producing those chunks of the target word for which they have established, well-practiced production routines; in this case, CV and VC chunks. Our findings thus have broader implications for the relationship between lexical storage and production grammars (see e.g. [1], [7]).

(1) Stimuli (examples)

Cluster size	Word position	Sonority
CC: **bn**apa, **db**ota, **lb**uka	Onset: **br**upa, **pfl**ata	Rising: **kr**ata, kabu**kr**
CCC: **gmr**ota, **ptk**oka, **nzb**oka	Coda: tabo**lk**, tako**dnl**	Falling: **lb**uka, tabo**lk**
		Level: **db**ota, paka**pk**

(2) Repairs in clusters

Repair to cluster	#tokens	Example target word	Pronunciation
Deletion	624	ptkoka	ptoka
Segmental change	436	patubml	patugmn
Epenthesis (C or V)	128	pmota mtupa	ptmota mutupa
Assimilation	129	mnota	n:ota
Metathesis	124	pakatp	pakapt

[1] Becker, M. & A.-M. Tessier. 2011. Trajectories of faithfulness in child-specific phonology. *Phonology* 28: 163–196.

[2] Bernhardt, B. H. & J. P. Stemberger. 1998. *Handbook of phonological development from the perspective of constraint-based nonlinear phonology*. San Diego: Academic Press.

[3] Gathercole, S. E., C. S. Willis., A. D. Baddeley & H. Emslie. 1994. The Children’s Test of Nonword Repetition: a test of phonological working memory. *Memory* 2: 103–127.

[4] Gerlach, S. 2010. The acquisition of consonant features sequences: Harmony, metathesis and deletion patterns in phonological development. Ph.D. dissertation, University of Minnesota.

[5] Kavitskaya, D., M. Babyonyshev, T. Walls, & E. Grigorenko. 2011. Investigating the effects of phonological memory and syllable complexity in Russian-speaking children with SLI. *Journal of Child Language* 38: 979–998.

[6] Marshall, C., S. Ebbels, J. Harris, and H. van der Lely. 2002. Investigating the impact of prosodic complexity on the speech of children with Specific Language Impairment. *UCL Working Papers in Linguistics* 14: 43–68.

[7] McAllister Byun, T., Inkelas, S., & Rose, Y. 2016. The A-map model: Articulatory reliability in child-specific phonology. Forthcoming in *Language*.

[8] Rodrigues, A. & D. M. Befi-Lopes. 2009. Phonological working memory and its relationship with language development in children. Pro-Fono 21: 63–68.

Gestural adaptation when 'broadening' an L2 accent: An exploratory EPG study

Alexei Kochetov & Kiranpreet Nara
University of Toronto

Understanding phonetic changes during the processes of bilingual code-switching and L2 acquisition is important as it can provide important insights into interactions between distinct bases of articulation and language-specific phonemic representations [1,2,3]. While considerable work has been done on the phonetics of bilingual speech, most research has focused on a handful of acoustic variables (e.g. VOT in stops [4,5]), with relatively few works exploring changes at the level of articulation [6,7,8,9]. This paper presents results of an exploratory study examining whether and how the production of coronal sounds is adapted when the speaker intentionally modifies their L1-influenced English accent – from their 'normal' pronunciation to an exaggerated, 'broad' accent. This is done using electropalatography (EPG, [10]) – the method that employs a custom-made artificial palate with built-in electrodes to track the contact between the tongue and the roof of the mouth.

EPG data were collected from a female Punjabi-Canadian English bilingual, who has been using English extensively since the age of 11. Her speech can be characterized as having a rather subtle non-native English accent. A 'broad' version of this accent is expected to manifest a fuller set of phonetic/phonological characteristics of the speaker's L1. Note that Punjabi employs a contrast between dentals and retroflexes, and the use of [ʈ, ɖ] instead of [t, d] is typical for Punjabi-influenced English [11] (and Indian English in general [12]). Punjabi also differs from English in the quality of the corresponding lateral and rhotic sounds (a clear /l/ and an alveolar trill/tap /r/), while lacking English-like dental fricatives /θ, ð/ (as shown in (1)). It was therefore expected that articulatory changes, if occurred, would primarily involve these most salient L1/L2 differences.

The materials for the study consisted of a set of English words with 12 coronal consonants in word-initial position (shown in 2a), produced in a carrier sentence (2b). The same list was recorded using the speaker's 'normal' English accent and using a 'broad', intentionally exaggerated Punjabi-accented English. Articulatory measurements were extracted at the point of maximum contact for each consonant constriction from 9 repetitions of the utterances (216 tokens in total). The dependent variables were (a) the *quotient of activation* over the entire palate (Q, the number of 'on' electrodes divided by the total number, 62) and the *centre of gravity* of contact along the front/back dimension (CoG) [10]. These were expected to distinguish between small-scale differences in constriction degree and location of consonants. T-tests were performed to compare the two conditions separately for each consonant.

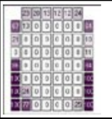
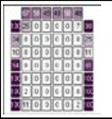
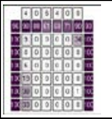
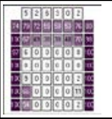
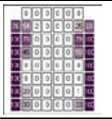
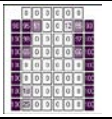
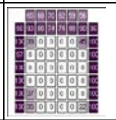
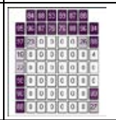
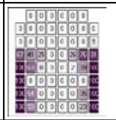
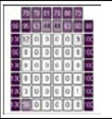
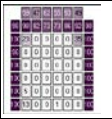
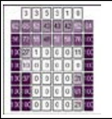
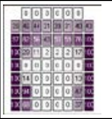
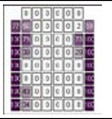
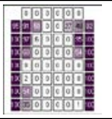
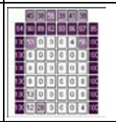
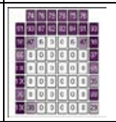
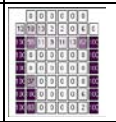
The results revealed that some consonants were significantly affected by the conditions, while others remained essentially the same. This can be seen in linguopalatal contact profiles in (3). Specifically, the tongue constrictions for /θ, ð/ were shifted backwards in the broad accent (BA) compared to the normal accent condition (NA), reflective of a change in place (dental to denti-alveolar). A similar place shift from alveolar to post-alveolar (retroflex) was observed for /t/; while not showing the retraction, /d/ exhibited a higher posterior side contact in BA, indicative of a retroflex-like raising. The lateral had more anterior side contact in BA, reflecting the raising/fronting of the tongue body (a clear /l/). The rhotic showed a more anterior constriction, indicative of a shift from the retroflex to the alveolar place. The other consonants (/s, z, n/ as well as /ʧ, ʤ, ʃ/ not shown in (3)) did not exhibit significant differences, and thus seem to have employed the same gestures. We are currently comparing the BA contact patterns to those exhibited by the same speaker in her L1 (Punjabi) productions.

While limited to a single-speaker case, the results provide an interesting insight into gestural adjustments made during the process of L2 accent manipulation. These findings can be used for further articulatory work investigating bilingual/L2 speech and phonetic imitation.

(1) English coronals: /θ, ð, t, d, ʧ, ʤ, s, z, ʃ, ʒ, n, l, ɹ/
Punjabi coronals: /t̪, t̪ʰ, d̪, ʈ, ʈʰ, ɖ, ʧ, ʧʰ, ʤ s, ʂ, n, ɳ, l, ɭ, r, ɽ/

(2) a. *thigh, thy, tie, die, chai, jive, sigh, ‘zye’, shy, nigh. lie, rye*
b. *He saw* ___ (repeated 3 times).

(3) Linguopalatal contact profiles for English coronal consonants produced by the speaker under the *normal accent (NA)* and *broad accent (BA)* conditions, with t-test results for Q and CoG measures indicated below (***$p<0.001$, **$p<0.01$, *$p<0.05$, *n.s.*=‘not significant’).

	/θ/	/ð/	/t/	/d/	/s/	/z/	/n/	/l/	/ɹ/	
NA										*front* *back*
BA										*front* *back*
Q	***	***	***	n.s.	n.s.	n.s.	n.s.	***	**	
CoG	***	*	*	*	n.s.	n.s.	n.s.	*	***	

References

[1] Flege, J. E. 1987. The production of “new” and “similar” phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics* 15, 47–65.

[2] Grosjean, F. 1998. Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition* 1, 131–49.

[3] Bullock, B. E. 2009. Phonetic reflexes of code-switching, In Bullock, B. E. & Toribio, A. J. (Eds.), *The Cambridge Handbook of Linguistic Code-switching*, Cambridge University Press, 163–81.

[4] Sancier, M. L., & Fowler, C. A. 1997. Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics* 25, 421–36.

[5] Olson, D. J. 2013. Bilingual language switching and selection at the phonetic level: Asymmetrical transfer in VOT production. *Journal of Phonetics* 41, 407–20.

[6] Nissen, S. L., Dromey, C., Wheeler, C. 2008. First and second language tongue movements in Spanish and Korean bilingual speakers. *Phonetica* 64, 201-216.

[7] Chakraborty, R., & Goffman, L. 2011. Production of lexical stress in non-native speakers of American English: Kinematic correlates of stress and transfer. *Journal of Speech and Language Hearing Research* 54, 821-835.

[8] Hacking, J. F., Smith, B. L., Nissen, S. L., & Allen, H. 2016. Russian palatalized and unpalatalized coda consonants: An electropalatographic and acoustic analysis of native speaker and L2 learner productions. *Journal of Phonetics* 54, 98–108.

[9] Ko, I. 2013. The articulation of Korean coronal obstruents: Data from heritage speakers and second language learners. Doctoral dissertation, University of Hawai’i at Manoa.

[10] Gibbon, F., & Nicolaidis, K. 1999. Palatography. In Hardcastle, W. & Hewlett, N. (Eds.), *Coarticulation: Data, theory and techniques*. Cambridge: Cambridge University Press. 229–45.

[11] Bhatia, T. 1993. *Punjabi: A Cognitive-descriptive Grammar.* London: Routledge.

[12] Gargesh, R. 2004. Indian English: Phonology. In Edgar, W., Burridge, K., Kortmann, B., Mesthrie, R., Upton, C. (Eds.), *A Handbook of Varieties of English, Vol 1: Phonology.* Berlin: Mouton de Gruyter, 993–1002.

Variability in realization of focal accent in Hungarian – articulatory and acoustic data

Alexandra Markó[1,2], Márton Bartók[1,2], Tamás Gábor Csapó[2,3], Andrea Deme[1,2] and Tekla Etelka Gráczi[2,4]

[1]*Eötvös Loránd University,* [2]*MTA-ELTE „Lendület" Lingual Articulation Research Group,* [3]*Budapest University of Technology and Economics,* [4]*Research Institute for Linguistics HAS*

While earlier studies focused predominantly on the role of f_0 in realization of focal accent in Hungarian, recent papers found that vowels occurring in accented syllables were significantly longer than their non-accented counterparts [1, 2, 3]. Several models (e.g., [4]) suggest that longer duration may lead to more accurate articulatory movements and thus the target might be better reached. Therefore, we may also assume that longer vowels in the accented position may be articulated with a greater effort (and with smaller variability, see [5]) in Hungarian, similarly to several other languages. However, as for Hungarian, there is an apparent consensus in the literature that vowel quality is not expected to co-vary with prominence (which is also a common pattern in several languages). Apart from a few earlier studies (see a review in [6]), which were largely inexplicit about the details of their methods, and a pilot study on a rather unbalanced material [3], acoustic correlates of vowel quality, i.e., formant structure, were not analysed reliably. Moreover, linguo-articulatory correlates of vowel quality in focal accent was not analysed with respect to Hungarian either. Nevertheless, recently a study revealed that in Hungarian pitch-accent has an effect on the variability of vowel both in the acoustic and in the articulatory domain [7].

In our study, utterance-initial vowels (/i u ɛ ɒ/) in preverbal focus vs. prefocal topic positions (both occurred sentence-initially) were compared with respect to their articulatory and acoustic parameters. Parallel acoustic and ultrasound recordings were made with 20 female speakers. With each participant, 40 target utterances (5 repetitions for each vowel in each condition) and 80 filler utterances were recorded. In a previous study carried out on the same material we found that vowel duration and the f0-peak alignment is different between the two conditions [8], therefore we may assume that focus bears higher prominence than topic. Since longer durations were observed in the focus condition we hypothesized that **formant values** and **tongue contours** differ as a function of condition. In accordance with [5, 7], we expected smaller variability in both measures in the focus condition.

F_1 and F_2 (mean, Hz) were automatically measured at the temporal midpoint of the vowel in Praat [9]. Formant frequencies were standardized within speakers using *z*-transformation [10] in the phonR package [11]. On the basis of F_1 and F_2 data, the Euclidean distance of the centroid of the vowel space and each token was also calculated [12]. Euclidean distances were compared using modified signed-likelihood ratio tests (MSLRTs) for equality of coefficient of variations [13, 14]. Tongue contours were manually traced on the ultrasound frame extracted from the temporal midpoint of the vowel, and variability of the tongue contours was measured by the Nearest Neighbour Distance (NND [15]) method. Linear mixed models were used to assess the effect of prominence and vowel quality on the measured variables.

Tongue contours showed smaller variability (NND) in the focus condition (Fig 1) for all vowels, but this tendency was not confirmed by statistical analysis. As for NND, smaller SDs were observed in focus, except for /u/. Statistically, acoustic vowel space (Fig 2) and formant frequencies did not appear to differ between the two conditions. However, while the variance of F_1 values did not differ significantly across conditions either, we found a significant difference in the variance of F_2 (MSLRT = 7.77, $p < 0.01$).

We assume that these patterns is in part due to the fact that some of the speakers tended to better reach the articulatory target in the focus condition than the others. Therefore, we plan to extend the above analysis looking in more depth into interspeaker variability.

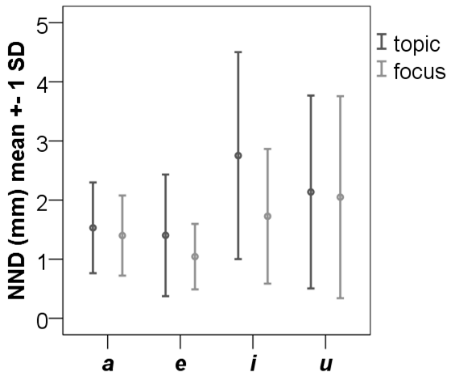


Figure 1. *Variability in tongue contour measured in NND (mm, mean±1 SD).*

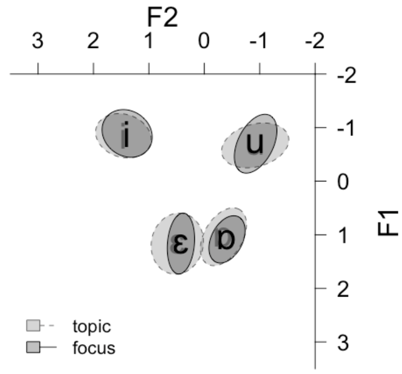


Figure 2. *Standardised F1 × F2 space of the analysed vowels as a function of condition*

[1] Szalontai, Á., Wagner, P., Mády, K., & Windmann, A. 2016. Teasing apart lexical stress and sentence accent in Hungarian and German. *Tagungsband 12. Tagung Phonetik und Phonologie im deutschsprachigen Raum (P&P 12)*, 216-219.

[2] Mády, K., Reichel, U., & Szalontai, Á. 2017. A prozódiai prominencia (nem)jelölése a németben és a magyarban. *Általános Nyelvészeti Tanulmányok* XXIX. 77–98.

[3] Markó, A., Bartók., M., Gráczi, T.E., Deme, A., & Csapó, T.G. 2018. Prominence effects on Hungarian vowels: A pilot study. *Proceedings of Speech Prosody 2018* (Poznan, Poland). https://www.isca-speech.org/archive/SpeechProsody_2018/pdfs/138.pdf.

[4] Lindblom, B. 1963. Spectrographic study of vowel reduction. *Journal of Acoustical Society of America* 35, 1773-1781.

[5] Farnetani E., Vagges, K., & Magno-Caldognetto E. 1985. Coarticulation in Italian /VtV/ sequences: A palatographic study. *Phonetica* 42, 78–99.

[6] Vértes O. A. 1982. A magyar beszédhangok akusztikai elemzésének kérdései. In Bolla, K. (ed.), *Fejezetek a magyar leíró hangtanból.* Budapest: Akadémiai Kiadó, 71-113.

[7] Deme, A., Bartók, M., Gráczi, T. E., Csapó, T. G., Markó, A. in press. The effect of pitch accent on V-to-V coarticulation induced acoustic and articulatory variability of vowels.

[8] Markó, A., Bartók, M., Csapó, T. G., Deme, A., Gráczi, T. E. in press. The effect of focal accent on vowels in Hungarian: Articulatory and acoustic data.

[9] Boersma, P., Weenink, D. 2018. Praat: doing phonetics by computer [Computer program]. Version 6.0.43, http://www.praat.org/

[10] Lobanov, B. M. 1971. Classification of Russian vowels spoken by different speakers. *J. Acoust. Soc. Am.* 49. (2), 606–608.

[11] McCloy, D. R. 2016. phonR: tools for phoneticians and phonologists. R package version 1.0-7.

[12] Bradlow, A. R., Torretta, G. M., Pisoni, D. B. 1996. Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication* 20, 255–272.

[13] Krishnamoorthy, K., Lee, M. 2014. Improved tests for the equality of normal coefficients of variation. *Computational Statistics* 29(1-2), 215–232. http://link.springer.com/article/10.1007/s00180-013-0445-2

[14] Marwick, B., Krishnamoorthy, K. 2018. cvequality: Tests for the Equality of Coefficients of Variation from Multiple Groups. R software package version 0.1.3. https://github.com/benmarwick/cvequality

[15] Zharkova, N., Hewlett, N. & Hardcastle, W.J. 2012. An ultrasound study of lingual coarticulation in /sV/ syllables produced by adults and typically developing children. *Journal of the International Phonetic Association* 42/2, 193-208.

Resistance to resyllabification in Yucatecan Spanish: effects of prosodic structure and contact dynamics

Nuria Martínez García, Francesco Cangemi and Martine Grice
University of Cologne, Germany

Spanish is reported to resyllabify across word boundaries. Evidence is drawn from many varieties, although resyllabification has been shown to be gradient to some extent both in its acoustic realization and in the production of speakers ([1], [2], [3], [4]). Resyllabification across word boundaries is a process by which the syllabic affiliation of a word-ending sound changes when followed by a word-initial vowel. Thus, a word-final consonant becomes the onset of the following syllable (see examples 1 and 2), whereas for word-final vocalic sequences, a new syllable is created, either by means of vowel deletion (*se hace* 'it is done' [se.ˈa.θe] > [ˈsa.θe]), coalescence (*nombre exacto* 'precise name' [ˈnom.bɾe.e.ˈsak.to] > ˈnom.bɾe.ˈsak.to]) or diphthongization (*fuego y* 'fire and'[ˈfwe.ɣo.i] > [ˈfwe.ɣoi̯]). However, Yucatecan Spanish seems to resist to it, maybe because of Yucatec Maya influence.

The current study focuses on the prosodic word, which is defined as the domain of word stress [5] in which unstressed function words merge with following content words [5], [6]. We analyze Yucatecan Spanish, a Mexican variety in contact with Yucatec Maya. Since Maya has both glottalized vowels and consonants [7], and Yucatecan Spanish is characterized by glottalization—thought to be rare in Spanish [8]—which can block resyllabification [9], this variety may show greater resistance to resyllabification than others, particularly for Spanish–Maya bilingual speakers. The study explores i) whether resistance to resyllabification differs within and across prosodic words, and ii) whether knowledge of Maya has an effect on said resistance.

Sequences with a potential for resyllabification were coded for its presence or absence by a native speaker of Spanish based on auditory inspection. They were also coded for glottalization based on visual inspection of the waveforms and spectrograms. The data analyzed consisted of 150 observations per speaker taken from a semi-spontaneous speech corpus of interviews about language and culture. (1) is an example with a potential for resyllabification within the prosodic word, whereas (2) is another example across prosodic words (with/without resyllabification).

(1) *en agua* [e.ˈna.ɣwa] / [en.ˈa.ɣwa] 'in water'

(2) *todos andaban* [ˈto.ðo.san.ˈda.βan] / [ˈto.ðos.an.ˈda.βan] 'all walked'

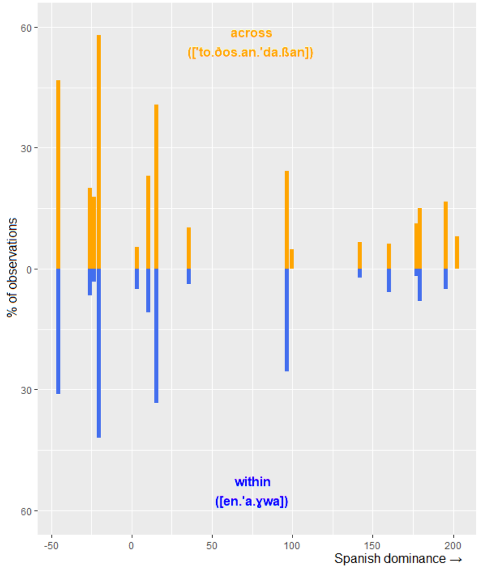
Speakers' language knowledge was assessed by means of the Bilingual Language Profile questionnaire [10], which yields a score about language dominance (in this case, Maya–Spanish). Sixteen participants (8 female, 8 male) received each a BLP score that placed them on a continuum from Maya dominance to only knowledge of Spanish. Just one speaker reported knowledge of another language (English). Prosodic position was coded as *within* the prosodic word and *across* prosodic words. Bayesian regression models with weakly informative priors were fitted using the `brms` package [11] in R [12], with resyllabification as the binomial dependent variable, prosodic position and BLP score (scaled) as independent variables, speaker as random intercept, and by-subject random slopes for the effect of prosodic position, and checked for a two-way interaction. Results (Table 1 and Figure 1) show evidence for an effect of prosodic position and knowledge of Maya. Resistance to resyllabification is greater across than within prosodic words (orange bars vs. blue bars), and it is also greater the more Maya-dominant the scores are (bars to the left are higher than bars to the right). There is no interaction between prosodic position and BLP score. 81.5 % of sequences perceived as resisting resyllabification presented cues to glottalization (creakiness, burst and/or silence).

To sum up, prosodic position and knowledge of Maya play a considerable role in resistance to resyllabification in Yucatecan Spanish, with the productions of all participants being affected by prosodic position and language knowledge.

Table 1. Main effects of prosodic position and BLP score (scaled) and their interaction.

	Estimate	Lower CrI	Upper CrI
Intercept	−2.60	−3.25	−1.98
pros_pos(across)	0.93	0.57	1.34
blp_score_s	−0.66	−1.26	−0.04
pros_pos(across):blp_score_s	0.11	−0.23	0.48

Figure 1. Resistance to resyllabification in the prosodic word domain. Each bar represents a speaker. Speakers are ordered from Maya dominance (left) to Spanish dominance (right).



[1] J. I. Hualde, M. Simonet, and F. Torreira, “Postlexical contraction of nonhigh vowels in Spanish,” *Lingua*, vol. 118, no. 12, pp. 1906–1925, 2008.

[2] P. Strycharczuk and M. Kohlberger, “Resyllabification reconsidered: On the durational properties of word-final /s/ in Spanish,” *Journal of the Association for Laboratory Phonology*, vol. 7, no. 1, pp. 1–24, 2016.

[3] L. Aguilar, “Los enlaces vocálicos: ¿Cuestión de dominios prosódicos?,” *Revista Internacional de Lingüística Iberoamericana*, vol. 3, no. 2, pp. 29–48, 2005.

[4] M. C. Alba, “Accounting for variability in the production of Spanish vowel sequences,” in *Selected Proceedings of the 9th Hispanic Linguistics Symposium*, N. Sagarra and A. J. Toribio, Eds. Somerville, MA: Cascadilla Proceedings Project, 2006, pp. 273–285.

[5] J. I. Hualde, “Stress removal and stress addition in Spanish,” *Journal of Portuguese Linguistics*, vol. 6, no. 1, pp. 59–89, 2007.

[6] J. I. Hualde, “Unstressed words in Spanish,” *Language Sciences*, vol. 31, no. 2–3, pp. 199–212, 2009.

[7] H. Avelino, E. Shin, and S. Tilsen, “The phonetics of laryngealization in Yucatec Maya,” in *New perspectives in Mayan linguistics*, H. Avelino, Ed. Newcastle upon Tyne: Cambridge Scholars Publishing, 2011, pp. 1–20.

[8] M. Garellek, “Voice quality strengthening and glottalization,” *Journal of Phonetics*, vol. 45, pp. 106–113, 2014.

[9] J. Michnowicz and L. Kagan, “On glottal stops in Yucatan Spanish,” in *Spanish language and sociolinguistic analysis*, S. Sessarego and F. Tejedo-Herrero, Eds. Amsterdam and Philadelphia: John Benjamins, 2016, pp. 217–240.

[10] D. Birdsong, L. M. Gertken, and M. Amengual, *Bilingual Language Profile: An easy-to-use instrument to assess bilingualism*. COERLL, University of Texas at Austin, 2012.

[11] P.-C. Bürkner, “brms: An R package for Bayesian multilevel models using Stan,” *Journal of Statistical Software*, vol. 80, no. 1, pp. 1–28, 2017.

[12] R Core Team, Vienna: R Foundation for Statistical Computing, 2018.

Disentangling metrical prominence from segmental and word-boundary effects

Janina Mołczanow[1], Beata Łukaszewicz[1] and Anna Łukaszewicz[1]
[1]*University of Warsaw*

This paper reports on the results of an acoustic study of the correlates of primary and secondary stress in Ukrainian. The experiment has been designed to control for the intrinsic differences in segmental length and influence of word boundaries, which can potentially constitute confounding factors in the studies of metrical prominence (e.g., [1], [2], [3]). Ukrainian has free lexical stress and predictable rhythmic stress ([4], [5], [6]). Lexical stress can appear on any syllable in the word and rhythmic stress can be located at both word edges, provided that at least one unstressed syllable intervenes between the rhythmic and lexical stress. This is illustrated in (1) with minimal pairs of quadrisyllabic words with lexical stress placed on the second and the third syllable.

In the present study, we performed pairwise comparisons of consonants and vowels in segmentally identical pairs of quadrisyllabic words differing in the position of stress, such as the minimal pairs exemplified in (1). The words contained only canonical CV syllables. Four metrical configurations were examined: pretonic - secondary (1st syllable), tonic - pretonic (2nd syllable), unstressed - tonic (3rd syllable), secondary – unstressed (4th syllable). (We adopt the terminology widely used in the literature on Slavic prosody, e.g. [7], where 'tonic' refers to primary lexical stress, while 'pretonic' points to positions immediately preceding main stress.) The stimuli consisted of 7 minimal pairs, uttered in 3 repetitions, and filler items, which were embedded in a frame. The data were collected from 8 monolingual native speakers in Western Ukraine. Segmentation was done manually, and eight duration measurements were taken for each token. Annotation and automated measurements were conducted in Praat [8]. A potential problem in extrinsic comparisons is variable tempo of speech. To control for this, duration of each vowel was expressed in proportion to the average duration of vowel segments in a word. In sum, measurements from 2688 segments (1344 consonants and 1344 vowels) were entered into statistical analyses (done in SPSS, v. 25). Linear mixed effects models were built for each position separately to test the effect of stress on vowel duration. Intercepts and slopes for speaker and item were included in the random structure. For vowels, significant results were obtained for all positions (see Fig. 1): 1st syllable (pretonic *vs.* secondary stress: $\beta = 0.1$, $SE = 0.02$, $t = 4.61$, $p < 0.01$), 2nd syllable (tonic *vs.* pretonic: $\beta = 0.79$, $SE = 0.05$, $t = 16.75$, $p < 0.001$), 3rd syllable (unstressed *vs.* tonic: $\beta = -0.96$, SE = 0.05, $t = -19.65$ $p < 0.001$), 4th syllable (secondary stress *vs.* unstressed: $\beta = 0.08$, $SE = 0.03$, $t = 2.36$, $p = 0.045$). For consonants, the effect of stress was significant only for the second and third position (tonic *vs.* pretonic and unstressed, respectively).

In sum, the results indicate that lexical stress is expressed by the increased duration of both the consonant and the vowel. The comparison of the initial syllables points to the presence of pretonic lengthening, which, contrary to what one might expect, is stronger than the lengthening induced by the presence of rhythmic stress. However, pretonic lengthening has also been reported in previous acoustic studies on Ukrainian stress [5], [9], which looked at words containing more than four syllables. The presence of the same effect in quadrisyllabic words reported here corroborates the intuition expressed in these studies that pretonic lengthening occurs consistently in pretonic position independently of the number of syllables between the lexical position and the edge of the word, and hence of the rhythmic structure of words. A significant difference in duration was observed in word-final position: the final vowel was longer in [σˈσσˌσ] than in [ˌσσˈσσ]. This result demonstrates that the acoustic prominence of the word-final vowel reported earlier ([4], [6]) is not conditioned by the presence of a word boundary, but is a realisation of secondary stress.

(1)	**σˈσ σˌ σ**	[pɔˈʋɔdɨˌtɨ] ‘to behave’	[dɔˈbʲiɦaˌtɨ] ‘to run, perf.’
	ˌσ σ ˈσ σ	[ˌpɔʋɔˈdɨtɨ] ‘to lead, perf.’	[ˌdɔbʲiˈɦatɨ] ‘to run, imperf.’

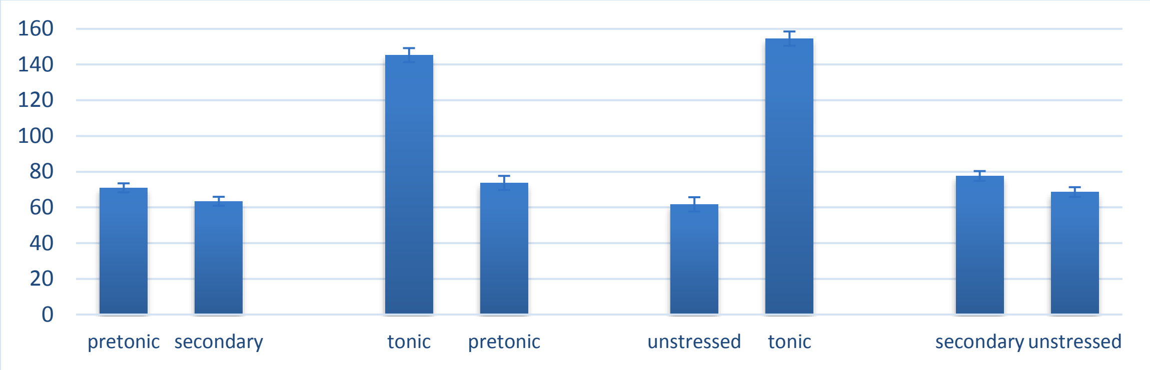


Figure 1. Mean vowel duration (ms) depending on position and stress. (Error bars: 95% confidence intervals adjusted for paired comparisons.)

[1] Plag, I., Kunter, G., & Schramm, M. 2011. Acoustic correlates of primary and secondary stress in North American English. *Journal of Phonetics* 39, 362-374.

[2] Turk, A. E., & Shattuck-Hufnagel, S. 2000. Word-boundary-related duration patterns in English. *Journal of Phonetics* 28, 397-440.

[3] Vogel, I., Athanasopoulou, A., & Pincus, N. 2016. Prominence, contrast and the functional load hypothesis: An acoustic investigation. In Heinz, J., Goedemans, R., & H. van der Hulst (Eds.), *Dimensions of Phonological Stress*. Cambridge: Cambridge University Press, 123-167.

[4] Nakonečnyj, M. F. 1969. Naholos. In Bilodid, I. K. (Ed.), *Sučasna ukrajins'ka literaturna mova. Vstup. Fonetyka*. Kyjiv: Naukova Dumka, 358-369.

[5] Łukaszewicz, B., & Mołczanow, J. 2018. Rhythmic stress in Ukrainian: Acoustic evidence of a bidirectional system. *Journal of Linguistics* 54, 367-388.

[6] Łukaszewicz, B., & Mołczanow, J. 2018. Leftward and rightward stress iteration in Ukrainian: acoustic evidence and theoretical implications. In Czaplicki, B., Łukaszewicz, B. & M. Opalińska (Eds.), *Phonology, Fieldwork and Generalisations*. Berlin: Peter Lang, 259-275.

[7] Bethin, C. Y. 2006. Stress and tone in East Slavic dialects. *Phonology* 23, 125–156.

[8] Boersma, P., & Weenink, D. 2016. *Praat: doing phonetics by computer*. Computer program. www.praat.org.

[9] Łukaszewicz, B., & Mołczanow, J. 2018. The role of vowel parameters in defining lexical and subsidiary stress in Ukrainian. *Poznań Studies in Contemporary Linguistics* 54, 355-375.

Phonetic encoding of phonological representation of hiatus in Romanian: a study of durational patterns

Oana Niculescu[1], Ioana Vasilescu[2], Martine Adda-Decker[2]
[1]"Iorgu Iordan – Al. Rosetti" Linguistic Institute, Romania
[2]LIMSI, CNRS, Université Paris-Saclay, Bât. 508, Orsay, France

Context. This study investigates duration patterns of hiatus with the aim of highlighting the differences between canonical (phonological) representations of VV-sequences vs. phonetic actualization in continuous speech as portrayed by two speaking styles in Romanian. It is commonly known that connected speech entails systematic variability with respect to various reduction processes ([1]), meaning that hiatus is shorter in spontaneous speech than in careful speech. In order to get insight into the acoustic patterns that underlie the different realizations of hiatus, we compared duration patterns of internal (IntH – within the word) and external hiatus (ExtH – across word-boundaries) in spontaneous (SS) and read speech (RS).

Data and methodology. In this presentation we focus on data of one Romanian speaker extracted from a larger database of nine subjects representative of the Southern dialect of Muntenia. Each participant had to perform two tasks. For the *monologue task*, the subjects were required to talk about their previous summer activities (approximately 40 min per speaker). For the *reading task*, all vocalic sequences of IntH and ExtH were extracted from the previous experiment, the tokens were placed in carrier-sentences, and three repetitions were elicited. The selected male speaker produced 40 min of SS, and 100 min of RS. The total number of vocalic pairs analyzed was 53 (1244 tokens), out of which 21 IntH (420 tokens), and 32 ExtH (824 tokens), showcasing different vocalic qualities. The recordings were forced aligned with an automatic speech transcription system described in [2] (annotated corpora have proven to be of high relevance in acoustic analysis, testing various linguistic hypothesis and exploring sound change and variation [3]; [4]). The automatic extraction was manually checked. We looked at the global duration of hiatus, measured from the onset of F1 in the first vowel until the offset of F2 in the second vowel. The temporal differences of VV-sequences in the two speaking styles were analysed according to stress position, distribution in the word, number of syllables, vowel height, and place of articulation (variables added manually to the forced aligned output) [5]. In this article we draw attention on the overall duration of IntH and ExtH as portrayed in different duration intervals.

Results. Previous work has focused mainly on hiatus-diphthong distinction ([6]; for a monographic outline of hiatus in Romanian, see [7]). The novelty of the present account consists in capturing the dynamics behind different repair mechanisms by proposing **5 duration intervals** (based on data distribution): *very short* (20–70ms), *short* (80–130ms), *medium* (140–220ms), *long* (230–340ms), and *very long* (350–500ms) see Figure 1. In this interpretation of avoidance strategies, (very) short duration intervals favor elision, while (very) long intervals highlight epenthesis, and possibly hesitation (especially for ExtH). Medium duration intervals can entail hiatus maintenance (when analyzing variation in vowel sequences in five Romance varieties [8], the authors concluded that Romanian presents the most robust hiatus-diphthong contrast). Our data show that ExtH is shorter than IntH ($p < 0.001$) independent of speech context. Moreover, based on number of intervals, we can infer that ExtH has a higher degree of variability than IntH, meaning it attracts more repair strategies (eg. ExtH /u.a/ can be avoided by epenthesis, diphthongization or elision – explaining why ExtH has a very short duration interval, while IntH /u.a/ is resolved by epenthesis and/or diphthongization). The results are summarized in Table 1. Due to unbalanced data and heterogeneity of variance, a Brown-Forsythe and Welch test were conducted, followed by a Games-Howel post-hoc test, showing that all levels of analysis are statistically representative. We conclude that there is a gradient acoustic continuum (i.e.

gradient phonetic outputs with respect to categorical phonological categories [8]) between hiatus in SS and RS. The results will be complemented by gathering additional data. This analysis opens up discussion in relation to modelling gradient phonetic and phonological distinctions between IntH and ExtH. By employing duration intervals we can gain a better understanding of the dynamics behind hiatus resolution strategies with valuable implications in language variation ([9]).

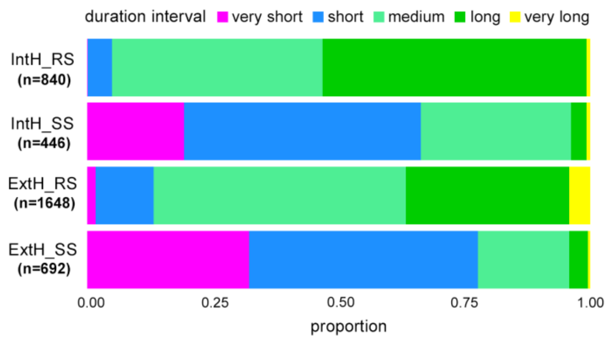


Figure 1. *Duration intervals of internal and external hiatus in spontaneous and read speech.*

	DURATION INTERVALS				
	very short	**short**	**medium**	**long**	**very long**
SS	27.2% *(n = 309)*	46.1% *(n = 525)*	22.7% *(n = 258)*	3.5% *(n = 40)*	0.5% *(n = 6)*
IntH in SS	19.3% *(n = 86)*	47.1% *(n = 210)*	29.8% *(n = 133)*	3.1% *(n = 14)*	0.7% *(n = 3)*
ExtH in SS	32.2% *(n = 223)*	45.5% *(n = 315)*	18.1% *(n = 125)*	3.8% *(n = 26)*	0.4% *(n = 3)*
RS	1.2% *(n = 29)*	9.3% *(n = 231)*	47.3% *(n = 1178)*	39.2 % *(n = 976)*	3% *(n = 74)*
ExtH in RS	1.7% *(n = 28)*	11.5% *(n = 190)*	50.2% *(n = 827)*	32.5% *(n = 535)*	4.1% *(n = 68)*
ExtH in SS	32.2% *(n = 223)*	45.5% *(n = 315)*	18.1% *(n = 125)*	3.8% *(n = 26)*	0.4% *(n = 3)*

Table 1. Statistics and n^{o} of occurrences of duration intervals in six contexts: (1) spontaneous speech, (2) internal hiatus in spontaneous speech, (3) external hiatus in spontaneous speech (in SS, IntH has less very short and short intervals than ExtH, and more medium duration intervals), (4) read speech, (5) internal hiatus in read speech, (6) external hiatus in read speech (in RS, IntH has less short intervals than ExtH, and more long intervals)

[1] Ernestus, M., Hanique, I., & Verboom, E. 2015. The effect of speech situation on the occurrence of reduced word pronunciation variant, *Journal of Phonetics*, 48, 60-75.
[2] Vasilescu, I.,Vieru, B., & Lamel, L. 2014. Exploring pronunciation variants for Romanian speech-to-text transcription, *Proceedings of SLTU-2014*, 161-168.
[3] Adda-Decker, M. 2006. De la reconnaissance automatique de la parole a l'analyse linguistique des corpus oraux, *Journaux d'Etude sur la Parole*, France.
[4] Vasilescu, I., Dutrey, C., & Lamel, L. 2015. Large scale data based investigations using speech technologies: the case of Romanian, *Proceedings of the 8th Conference on Speech Technology and Human-Computer Dialogue "SpeD 2015"*, Bucharest, October 14-17, 6 p.
[5] Niculescu, O., Vasilescu, I., & Adda-Decker, M. Duration Patterns of Internal and External Hiatus as a Function of Speaking Style, *Revue Roumaine de Linguistique*, to appear.
[6] Chitoran, I., & Hualde, J. I. 2007. From hiatus to diphthong: the evolution of vowel sequences in Romance, *Phonology*, 24, 37-75.
[7] Niculescu, O. 2018. *Hiatul intern şi hiatul extern în limba română contemporană. O analiză acustică* [Internal and external hiatus in contemporary standard Romanian. An acoustic analysis], Ph.D. dissertation, University of Bucharest.
[8] Boersma, P. 1998. *Functional Phonology: Formalizing the Interactions Between Articulatory and Perceptual drives*. Ph.D. dissertation, University of Amsterdam.
[9] Ohala, J.J. 1989. Sound change is drawn from a pool of synchronic variation. In: Breivik, LE, Jahr, EH (eds.) *Language change: Contributions to the study of its causes*. Mouton de Gruyter,173-198.

The yes-no question contour in Northern Brazilian Portuguese: revisiting the geographical continuum

Priscila Santos[1], Sónia Frota[2] and Marisa Cruz[2]
[1]Federal University of Rio de Janeiro, [2]University of Lisbon

According to previous studies, the intonational system of Portuguese has been characterized by the prevalence of bitonal pitch accents and monotonal boundaries ([1], [2]). In Brazilian Portuguese, only the complex boundary tone HL% was reported, in the rising-falling yes-no questions characteristic of the South (Santa Catarina and Rio Grande do Sul). The distribution of the nuclear contour of yes-no questions along the Atlantic Coast has provided evidence for a geographical continuum, with the Northeast presenting a rising contour, and the Center and South a rising-falling contour ([3], [4]). In the Center regions, unlike in the South, the nuclear contour exhibits a high tone clearly associated to the pitch accent, followed by a low boundary tone. These findings lead to the hypotheses that (i) yes-no questions in the extreme North of Brazil are mostly produced with a rising contour, according to the linguistic continuum previously described, and that (ii) a monotonal boundary predominates in the Northern Brazilian Portuguese yes-no questions.

To test these hypotheses, a semi-spontaneous eliciting task was used to collect data in Oiapoque, in the extreme North of Brazil (Fig.1), consisting in a fixed carrier sentence ending with nuclear words with different stress patterns – oxytone, paroxytone and proparoxytone (20, 30 and 20 sentences, respectively) –, that were presented to participants in individual written cards. Following the Autosegmental Metrical approach ([5], [6], i.a.), and using the P-ToBI system ([7]), data were analysed in Praat ([8]). Due to the availability of more segmental material (thus allowing a full realization of the intonational contour), yes-no questions with proparoxytone nuclear words were selected for the present analysis.

Preliminary results from two speakers show that a rising-falling nuclear contour is consistently used to produce yes-no questions, instead of the expected rising contour that prevails in the Northeast, according to the proposal of a geographical continuum. A closer inspection of the tonal boundary type shows that the simple low boundary tone prevails (62%) over the complex falling boundary tone (38%), which is, however, also possible (Fig.2). Given the predominance of the LHL melody in the extreme North of Brazil, and in order to provide additional acoustic evidence for the simple *versus* complex boundary tone, the peak height attained in the stressed and post-stressed syllables was manually extracted in semitones. Considering the range of 3 semitones as the reference for the peak height to be recognized as a H boundary tone ([9]), and focusing on the post-tonic stretch, we compared the peak height of the first and second post-stressed syllables (respectively, PT1 and PT2). A simple (L) boundary tone is annotated when the peak height of PT1 is 3 semitones higher than PT2; inversely, a complex (HL) boundary tone is annotated when the peak height difference between PT1 and PT2 is lower than 3 semitones. In our data, the simple boundary tone was reflected in a peak height difference between PT1 and PT2 ranging from 3 to 5 semitones (av. 4st), whereas the complex boundary tone was reflected in a peak height difference of 1-2 semitones between PT1 and PT2. Thus, peak height information in the post-tonic stretch adds to tonal alignment differences in characterizing simple and complex boundary tones.

In sum, preliminary findings from Oiapoque, a region in the extreme North of Brazil that was previously unstudied, confirm the predominance of monotonal boundary tones in Portuguese ([1], [2]), while showing that the complex HL% found in South also appears in the North. Moreover, the proposal of a geographical continuum ([3], [4]) for the Atlantic Coast cannot be generalized to the extreme North of Brazil, thus suggesting that the geographical distribution of yes-no question nuclear contours, similarly to findings for

European Portuguese ([10]), is discontinuous in the Brazilian variety of Portuguese. Implications of the latter finding for possible language contact phenomena, given the border status of the extreme North, need to be explored.

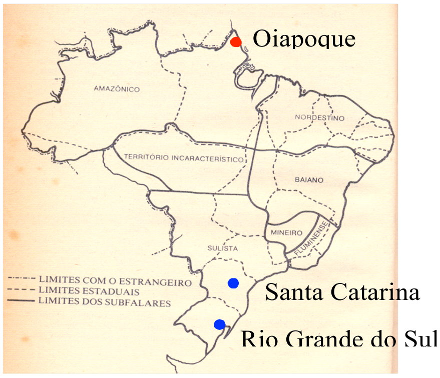


Figure 1. Data collection point (red bullet) in the North of Brazil, in the opposite vector of the already known data points (blue bullets). The map shows the main linguistic areas in Brazil, defined by Nascentes (1953).

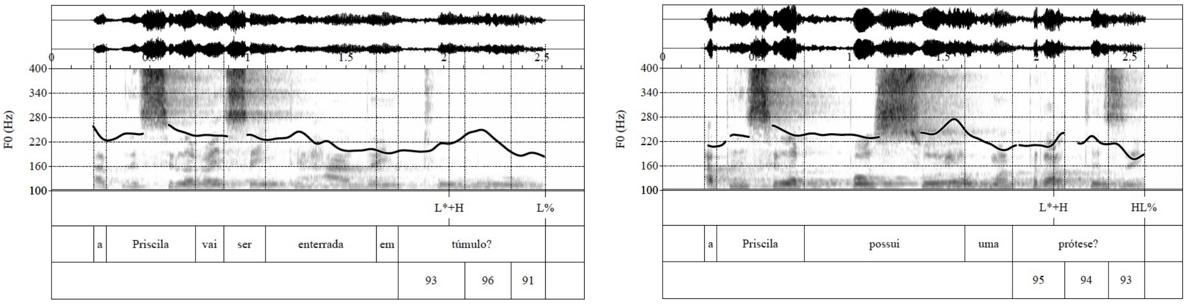


Figure 2. Rising-falling nuclear contour produced in yes-no questions in the North of Brazil: prevalence of the monotonal boundary tone (left panel – 'A Priscila vai ser enterrada em túmulo?' *Is Priscila going to be buried in a tomb?)* over the complex boundary tone (right panel – 'A Priscila possui uma prótese ?' *Does Priscila have a prosthesis?),* also possible. Values in the third tier refer to the peak height measured in semitones (re 1Hz).

[1] Frota, S., Cruz, M., Fernandes-Svartman, F., Collischonn, G., Fonseca, A., Serra, C., Oliveira, P., & Vigário, M. 2015. Intonational variation in Portuguese: European and Brazilian varieties. In S. Frota & P. Prieto (Eds), *Intonation in Romance* (pp. 235-283). Oxford: OUP.

[2] Frota, S., & Moraes, J. 2016. Intonation of European and Brazilian Portuguese. In W. L. Wetzels, S. Menuzzi & J. Costa (Eds.), *The Handbook of Portuguese Linguistics* (pp. 141-166). John Wiley & Sons, Inc.

[3] Castelo, J. 2016. *A entoação dos enunciados declarativos e interrogativos no Português do Brasil: uma análise fonológica em sete variedades da Costa Atlântica*. Unpublished PhD Dissertation. University of Lisbon.

[4] Castelo, J., & Frota, S. 2017. The yes-no question contour in Brazilian Portuguese: a geographical *continuum.* In P. Barbosa, M. C. Paiva & C. Rodrigues (Eds.), *Studies on Variation in Portuguese* (pp. 111-133). Amsterdam/Philadelphia: John Benjamins.

[5] Pierrehumbert, J. 1980. *The phonology and phonetics of English Intonation.* PhD Dissertation. Massachusetts, Institute of Technology.

[6] Ladd, R. 2008. *Intonational Phonology*. 2ª ed. Cambridge: CUP.

[7] Frota, S., Oliveira, P., Cruz, M. & Vigário, M. 2015. *P-ToBI: tools for the transcription of Portuguese prosody*. Lisboa: Laboratório de Fonética, CLUL/FLUL. http://labfon.letras.ulisboa.pt/InAPoP/P-ToBI/

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

[9] Falé, I., & Faria, I. 2006. Categorical perception of intonational contrasts in European Portuguese. In *Speech Prosody 2006*, Dresden, Germany.

[10] Cruz, M., Oliveira, P, Palma, P., Neto, B., & Frota, S. 2017. Building a prosodic profile of European Portuguese varieties: The challenge of mapping intonation and rhythm. In P. Barbosa, M. C. Paiva & C. Rodrigues (Eds.), *Studies on Variation in Portuguese* (pp. 81-110). Amsterdam/Philadelphia: John Benjamins.

A prosodic study of rhetorical questions in Italian

Patrizia Sorianello
University of Bari (Italy)

The present research investigates the prosodic realization of Italian questions having both literal and rhetorical meaning The rhetorical questions (henceforth RQs) are particular utterances whose syntactic structure, that of an interrogative, does not match their pragmatic meaning, that of an emphatic assertion. Unlike information-seeking questions (henceforth ISQs), RQs do not require an explicit answer because the addressee already knows it. They are indirect speech acts which transmit an obvious and presupposed meaning that goes beyond their literal interpretation [1], [2], [3], [4], [5].

In many languages, including Italian, RQs may have the same syntactic structure of a genuine question; in these cases, the contextual information, together with an adequate intonation contour, contribute to disambiguate them. This seems to suggest that the prosodic organization of RQs is to some extent different from that of ISQs [4]. Studies conducted in this direction show conflicting results [6], [7]. In general, RQs are more often characterized by a falling contour and a longer duration than ISQs [8], [9].

In Italian language the intonation of RQs did not receive sufficient attention so far: the observations are few and mostly obtained from impressionistic evaluations [10], [11]; [12].

This study aimed to analyze the intonational contour of rhetorical questions as compared to information-seeking questions. To this order, a corpus of 20 identical utterances (wh- and yes\no questions) was included in short dialogical scenarios and realized in their double pragmatic interpretation: as rhetorical and as literally meaning; for example, the following question *can you drive?* may have a literal interpretation, or changing the contextual frame, a rhetorical one. Ten native speakers of Bari Italian, aged 22-30 (5 males and 5 females) participated in this production experiment. They were asked to read the dialogical scenarios adopting a spontaneous intonation. Questions produced with particular emphasis, surprise or disappointment were considered separately.

The target sentences, (400 utterances: 200 RQs and 200 ISQs) were phonetically and phonologically analyzed using Praat. The following parameters were considered: f0 (average, maximum and minimum) pitch range, overall duration, nuclear vowel duration, speech rate, intensity. The intonation contour of each question was annotated by means of ToBI. We focused on the distribution of both nuclear pitch accent and boundary tone patterns. Our primary goal was to see whether f0 and duration contribute to the differentiation between RQs and ISQs. We hypothesize that the specific pragmatic functions of RQs are somehow reflected in their prosodic structure. We also predict longer duration in RQS than RQs.

The results of this analysis show that between RQs and ISQs there are prosodic differences, specially for wh- questions, the only ones to present the PA L* (29%), a wider pitch range and a longer nuclear vowel. RQs From the phonological perspective, boundary tones and nuclear pitch accents seem to have a different weight in the characterization of the RQs: in polar interrogatives the nuclear PA fails to distinguish between the two types of questions, being L+H* the most frequent nuclear PA in both RQs (93%) and ISQs (89%). A different behaviour concerns the realization of the boundary tones: in the most cases RQs are produced with low boundary tone L% (70%).

Besides intonation, duration plays a significant role: in RQs the stressed vowel that carries the nuclear pitch accent is always longer with respect to that of ISQs.

[1] Sadock, J.M. 1971. Queclaratives. *Papers from the Seventh Regional Meeting of the Chicago Linguistics Society*, (CLS 7), Chicago, Linguistics Society, 223-231.
[2] Frank, J. 1990. You call that a rhetorical question? Forms and functions of rhetorical questions in conversation, *Journal of Pragmatics* 14 (5): 723-738.
[3] Gutiérrez Rexach, J. 1998. Rhetorical question, relevance and scales, *Revista Alicantina de Estudios Ingleses* 11: 139-155.
[4] Han, C. 2002. Interpreting interrogatives as rhetorical questions, *Lingua* 112: 112-229.
[5] Rohde, H. 2006. Rhetorical questions as redundant interrogatives, *San Diego Linguistics Paper* 2, University of California, San Diego: 134-168.
[6] Miura, I. & Hara, N. 1995. Production and perception of rhetorical questions in Osaka Japanese, *Journal of Phonetics* 23: 291-303.
[7] Banuazizi, A. & Creswell, C. 1999. Is that a real question?: Final rises, final falls and discourse function in yes-no question intonation, *Proceedings of the 35th Regional Meeting of the Chicago Linguistics Society (CLS 35)*, Chicago, Linguistic Society: 1-14
[8] Bartels, S.C. 2013. *The intonation of English statements and questions. A compositional interpretation*, Routledge, New York & London, 1 ediz 1999, Garland Publishing.
[9] Wochner, D., Schlegel, J., Dehé, N. & Braun, B. 2015. The prosodic marking of rhetorical questions in German, *16th Annual Conference of the International Speech Communication Association, Interspeech* 2015: 987-991.
[10] Crisari, M. 1974. Sugli usi non istituzionali delle domande, *Lingua e Stile* 1: 29-56.
[11] Anzilotti, G.I. 1982. The rhetorical question as an indirect speech device in English and Italian, *Canadian Modern Language Review* 38: 290-302.
[12] Stati S. 1982. Le frasi interrogative retoriche, *Lingua e Stile* XVII(2): 195-207.

Investigating articulatory variability across recording sessions using a functional clustering technique

Lorenzo Spreafico[1], Alessandro Vietti[2], Alessia Pini[3], Simone Vantini[4]
[1]*Università degli Studi di Bergamo,* [2]*Libera Università di Bolzano,* [3]*Università Cattolica del Sacro Cuore,* [4]*Politecnico di Milano*

The ultrasound tongue imaging technique [1] is used to study the spatial and temporal properties of lingual articulation. Over the years, there has been increasing standardization in the statistical analysis of tongue profiles from ultrasonographic data [2, 3, 4]. Despite this, the current methods do not yet offer effective solutions for comparing data across recording sessions or from different speakers. In this paper, we show how a non-parametric clustering technique can accommodate for possible misalignment of functional data and therefore how it can be used to investigate intraspeaker articulatory variability across recording sessions. This technique allows the simultaneous co-registration and clustering of curves produced in different recording sessions via a functional k-medoid alignment [5, 6], which represents an improvement over the analysis based on k-mean alignment presented in [7] as it provides cluster centroids which are more faithful to the elicited data.

In order to test the potential of the technique, we ran a pilot experiment. First, we devised a list of reduplicated $C_1V_1.C_1V_1$ Italian words and non-words in which vowels were restricted to /i, e, ɛ, a, ɔ, o, u/ and consonants to /p, t, k, s, ʃ/ (e.g.: [ˈpaːpa, ˈteːte, ˈʃɔːʃɔ]). Then we collected ultrasound data using a SonixTablet machine controlled from a PC running Articulate Assistant Advanced software [8]. Data was recorded at about 100fps with 69 beam-formed echo pulses evenly spread over a 134-degree field of view. Data presented here are from the recordings of a trained adult male phonetician who read the randomized list of stimuli twenty times while fitted with the UltraFitty headset [9] to stabilize the ultrasound probe. After ten repetitions of the stimuli list, the headset was repositioned to simulate two distinct recording sessions.

Collected data were used to validate the clustering technique and test whether it was able to detect that the tongue shape was different or similar for:

(i) the articulation of the same vowel in the same phonetic conditions but across different recording sessions, e.g. variants of [a] by the same speaker in recording $session_1$ and recording $session_2$;

(ii) the articulation of the same vowel under different conditions within a recording session, e.g. positional variants of [a] in [pa], [ta], [ka], [sa], [ʃa].

To this aim, for each pair of vowels in the dataset we ran a 2-, 3-, 4- and 5-clusters partition. We adopted the L2 distance between curves to compute a dissimilarity index of curves-medoids within each cluster, and affinities - namely shift and dilation - as the group of admissible warping functions. We used the index to evaluate the alignment/clustering of the curves to their medoid templates. Results confirm the quality of the clustering technique for comparing data across recording sessions. For example, Figure 1 illustrates how the tongue profiles for [a] and [e] that were recorded in distinct sessions are correctly reported to two homogeneous clusters. However, Figure 2 firstly shows that by increasing the number of clusters, the quality of classification does not markedly improve. This evidence suggests that fine-graded articulatory variability that is present in the original data is not captured, namely positional variables are not identified. Secondly, the reduction in the dissimilarity between the case without and with alignment supports the importance of managing the possible curve misalignment before performing any statistical analysis, either while collecting data - possibly by resorting to a bite plane trace that provides a consistent reference to allow images to be rotated and translated [10] -, or while pre-processing data for the statistical analysis.

During the presentation, we will discuss how this can depend both on the clustering technique and on the accuracy limits of the ultrasound tongue imaging technique, and we will give information regarding the optimum settings for obtaining the best results.

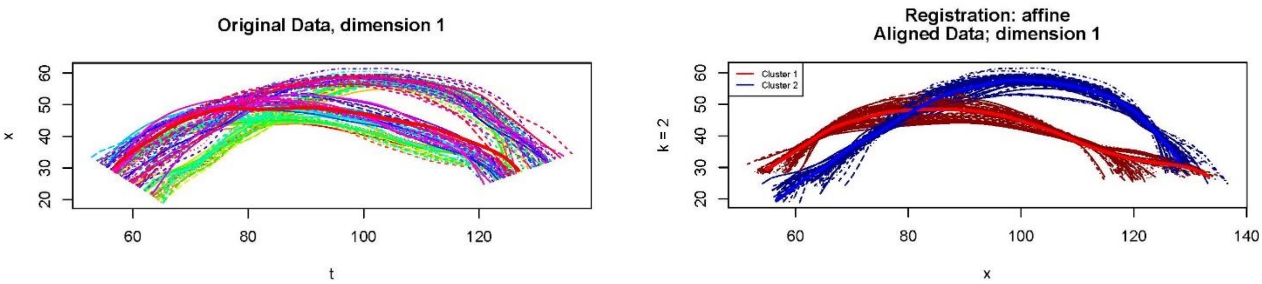


Figure 1. *Original curves (left) and aligned and clustered curves via 2-medoid alignment with affine warping (right). The colors in the left panel refer to the different positional variants of* [a] *and* [e].

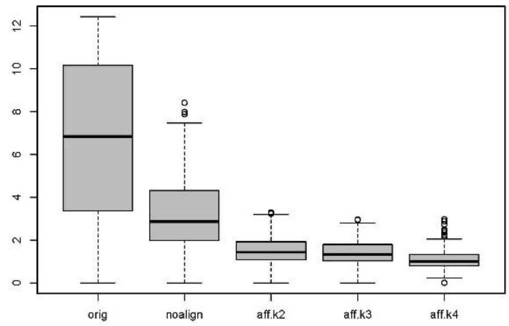


Figure 2. *Boxplot dissimilarity indexes for original, non-aligned, 2-, 3-, and 4-clusters partition for* [a] *and* [e].

[1] Stone, M. A guide to analysing tongue motion from ultrasound images. 2005. *Clinical linguistics & phonetics* 19, 455-501.

[2] Davidson, L. Comparing tongue shapes from ultrasound imaging using smoothing/spline analysis of variance. 2006. *The Journal of the Acoustical Society of America*, 120, 407-415.

[3] Dawson, K. M., Tiede, M. K., & Whalen, D. H. Methods for quantifying tongue shape and complexity using ultrasound imaging. 2016. *Clinical linguistics & phonetics*, 30, 328-344.

[4] Zharkova, N. Using ultrasound to quantify tongue shape and movement characteristics. 2013. *The Cleft Palate-Craniofacial Journal*, 50, 76-81.

[5] Sangalli, L. M., Secchi, P., Vantini, S., Vitelli, V. k-mean Alignment for Curve Clustering. 2010. *Computational Statistics and Data Analysis*, 54, 1219-1233.

[6] Sangalli, L. M., Secchi, P., Vantini, S., Vitelli, V. Functional clustering and alignment methods with applications. 2010. *Communications in Applied and Industrial Mathematics*, 1, 205-224.

[7] Vietti, A. Pini, A, Spreafico, L., Vantini, S. Towards cross-subject comparison of tongue profiles. A Functional Data Analysis approach. 2017. *Oral presentation at Ultrafest VIII*, Potsdam 4-6 October 2017.

[8] Articulate Instruments. 2011. *Articulate Assistant Advanced user guide*. Version 2.17.

[9] Spreafico, L., Pucher, M., Matosova, A. UltraFit: A Speaker-friendly Headset for Ultrasound Recordings in Speech Science. *Proceedings of Interspeech* 2018, 1517-1520.

[10] Scobbie, J., Lawson, E., Cowen, S., Cleland, J., Wrench, A. A common coordinate system for mid-sagittal articulatory measurement. 2010. *Working Paper WP-20*. QMU Edinburgh.

Acoustic evidence of /s/-retraction in the Northern Veneto dialect: a case-study

Ottavia Tordini, Vincenzo Galatà, Cinzia Avesani, Mario Vayra

1. The purpose of this contribution is to acoustically describe the socio-phonetic correlates of articulatory retraction involving the alveolar fricative /s/ as uttered by native speakers of Northern Veneto dialects (i.e. vernaculars; henceforth NVen) (for a classification of Veneto dialects, see e.g. [1]). Literature commonly defines /s/-retraction as a cross-linguistic phenomenon by which the alveolar fricative /s/ undergoes a retraction in its place of articulation, hence resulting as perceptually similar to the postalveolar fricative [ʃ] ([2, 3, 4, 5]). At acoustic level, evidence of such articulatory retraction lies in the spectral properties of the frication noise, with retracted-/s/ showing more energy in the low frequency regions than expected for a non-retracted /s/, resulting in lower values of Center of Gravity ([6, 7]). Conditions triggering /s/-retraction have been identified in specific prosodic contexts ([8, 9]), as well as in dialectal and social factors ([10]); moreover, [11] have recently investigated this phenomenon within the broader perspective of sound change occurring in some varieties of English.

2. Across Italy, evidence of this phenomenon is encountered in some dialectal and regional varieties (e.g., [12]); so far, however, motivations behind /s/-retraction have been explored in a single study on Cagliari Sardinian ([13]), essentially at sociolinguistic and stylistic level. Concerning Veneto varieties, the perceptual similarity between [s] and [ʃ] has been so far identified solely through auditory descriptions in [14] for Veneto Regional Italian (VRI), and acoustically explored in [15]. It is in fact acknowledged that while /s/ is produced in Standard Italian (SI) as a dental fricative ([16]) (or as a lamino-alveolar fricative in [17]), in VRI it is uttered as apical-alveolar, namely with a more back place of articulation than in SI, and as a non-laminal sound (it is labelled "*scibilante*" in [14]). However, to our knowledge, no specific acoustic analyses have been so far carried out to detect evidence of /s/-retraction among Veneto dialectal sub-systems. In light of these observations, our pilot study tries to fill the lack of experimental studies on this phenomenon. It is also worth noting that we are proposing for the first time an acoustic description of NVen native dialectal features, which from this perspective have so far been largely understudied (for an acoustic study on Veneto dialect as heritage language see [19]).

3. To empirically assess whether /s/-retraction takes place in NVen, we analyzed elicited spoken utterances of two male and two female adult NVen native speakers from the areas of Feltre and Cadore. We performed a narrow phonetic transcription through PRAAT ([20]), following the method already adopted in [15] and in [19]. Consistently with previous works on /s/-retraction ([6, 7]) which proved the effectiveness of spectral moment analyses for these purposes, we acoustically measured duration and four spectral moments (M1=Center of Gravity, M2=Standard Deviation, M3=Skewness, M4=Kurtosis) ([8, 9]) to classify the dialectal productions of target fricatives and to evaluate their acoustic nature. We retained approximately 500 tokens in VCV and CCV contexts; we compared the frication noise of **[s]** with the frication noise of the fricative release of the affricate **[tʃ]**. This choice is motivated by the absence of [ʃ] in the phonological inventory of all Veneto dialects ([1]), while it is present in the phonological system of the variety of Italian as spoken in the Veneto region.

4. Statistical analyses were subsequently run on acoustic data by fitting Linear Mixed Models with M1, M2, M3, M4 as dependent variables, and gender, target consonant, dialectal sub-variety (Feltrino vs Cadorino) as predictors. Post-hoc Tukey tests show that the sibilant fricatives [s] and the fricative release of [tʃ] do not differ in any spectral moment. Overall, preliminary data suggest an acoustic convergence of /s/ towards the fricative release of [tʃ] both in the spectral moments that index place of articulation (M1 and M3) and in those that index degree of laminality (M2 and M4). We are currently comparing the data of NVen with new data of another Veneto dialect, Central Veneto, to assess the extension of the phenomenon. For both dialects, we will further check for possible effects induced by the prosodic context, i.e whether /s/-retraction is affected by the stress status of the syllable in which it occurs.

References

[1] Zamboni, A. 1988. Italienisch: Areallinguistik IV. Venezien. In Holtus, G., Metzeltin, M., & Schmitt, C. (Eds), *Lexicon der Romanistischen Linguistik*. Tubingen: Niemeyer Verlag, 517-538.
[2] Shapiro, M. 1995. A case of distant assimilation: /str/→/Str/. *American Speech,* 70(1), 101-107.
[3] Mielke, J., Baker, A. & Archangeli, D. 2010. Variability and homogeneity in American English /ɹ/ allophony and /s/ retraction. In Kühnert, B. (Ed.), *Variation, detail, and representation. LabPhon 10*. Berlin: Mouton de Gruyter, 699-729.
[4] Baker, A., Archangeli, D. & Mielke, J. 2011. Variability in American English s- retraction suggests a solution to the actuation problem. In *Language variation and change*, 23, 347-374.
[5] Gylfadottir, D. 2015. Shtreets of Philadelphia: An acoustic study of /str/-retraction in a naturalistic speech corpus. *University of Pennsylvania Working Papers in Linguistics*, 21, 2, 2-11.
[6] Jongman, A., Wayland, R. & Wong, S. 2000. Acoustic characteristics of English fricatives. *Journal of the Acoustical Society of America,* 108, 1252-1263.
[7] Li F., Edwards, J. & Beckman, M. 2009. Contrast and covert contrast: The phonetic development of voiceless sibilant fricatives in English and Japanese toddlers. *Journal of Phonetics*, 37, 111-124.
[8] Bukmaier, V., Harrington, J. & Kleber, F. 2014. An analysis of post-vocalic /s- / neutralization in Augsburg German: evidence for a gradient sound change. *Frontiers in Psychology*, 5, 1-12.
[9] Phillips, J. B. 2016. Prosodic paths to sound change: effects of phrasal boundaries on /s/-retraction in American English. 21st Mid-Continental Phonetics & Phonology Conference, Michigan State University.
[10] Stuart-Smith, J., Sonderegger, M., Mcauliffe, M., Mcdonald, R., Mielke, J., Thomas, E., & Dodsworth, R. 2018. Dialectal and social factors affect the phonetic bases of English /s/-retraction. Poster presented at the 16 Conference on Laboratory Phonology - LabPhon16 (Lisbon).
[11] Stevens, M. & Harrington, J. (2016). The phonetic origins of /s/-retraction: Acoustic and perceptual evidence from Australian English. *Journal of Phonetics*, 58, 118-134.
[12] Crocco, C. 2016. Everyone has an accent. Standard Italian and regional pronunciation. In Cerruti, M., Crocco, C. & Marzo, S. (Eds.), *Towards a New Standard: Theoretical and Empirical Studies on the Restandardization of Italian*. Boston/Berlin: Mouton de Gruyter, 89-117.
[13] Mereu, D. 2017. Arretramento di /s/ nel sardo cagliaritano: uno studio sociofonetico. In Bertini C., Celata C., Lenoci, G., Meluzzi, C., & Ricci, I. (Eds.), *Sources and functions of speech variation. Disentangling the role of biological and social factors.* Milano: Officinaventuno, 45-65.
[14] Canepari, L. 1984. *Lingua italiana nel Veneto*. Padova: Clesp.
[15] Avesani, C., Galatà, V., Vayra, M., Best, C., Di Biase, B. & Ardolino, F. 2017. Phonetic details of coronal consonants in the Italian spoken by Italian-Australians from two areas of Veneto. In Bertini C., Celata C., Lenoci, G., Meluzzi, C., Ricci, I. (Eds.). *Sources and functions of speech variation. Disentangling the role of biological and social factors.* Milano: Officinaventuno, 281-306.
[16] Cortelazzo, M. (1972). *Avviamento critico allo studio della dialettologia italiana, Vol. III: Lineamenti di italiano popolare*. Pisa: Pacini.
[17] Loporcaro, M. & Bertinetto, P. M. (2005). The sound pattern of Standard Italian, as compared with the varieties spoken in Florence, Milan and Rome. *Journal of the International Phonetic Association,* 35(2),131-151.
[18] Boersma, P., & Weenink, D. 2016. *Praat: doing phonetics by computer*. Computer program.
[19] Avesani, C., Galatà, V., Vayra, M., Best, C.T., Di Biase, B., Tordini, O. & Tisato, G. (2015). Italian roots in Australian soil: coronal obstruents in native dialect speech of Italian-Australians from two areas of Veneto. In Vayra, M., Avesani, C. & Tamburini, F. (Eds.), *Il farsi e disfarsi del linguaggio. Acquisizione, mutamento e destrutturazione della struttura sonora del linguaggio*. Milano: Officinaventuno, 61-86.

Oral session 7

L2 learning

Prosodic effects on L2 French vowels: a corpus-based investigation

Fabián Santiago[1] and Paolo Mairano[2]

[1]*University of Paris 8 (SFL, LPP, LLF),* [2] *University of Lille (STL)*

Vowels in strong prosodic positions are produced with increased articulatory effort and expand farther apart within the vowel space. Such *prosodic strengthening* is correlated to the level of the prosodic hierarchy [2]. For instance, in French, the higher the prosodic domain (Intonation Phrases (IP) vs Accentual Phrases (AP)), the more vowels are acoustically expanded and hyper-articulated [4, 5]. Similarly, in English, vowels produced at IP-initial positions are more dispersed into the vowel space than those in accented positions [1]. These studies claim that vowels in strong prosodic positions are more *canonical*, since speakers enhance certain phonological features. Yet, it is unclear whether prosodic strengthening occurs similarly in other languages: according to [8, 9] the presence of pitch accents and/or lexical stress is not a good predictor of acoustic dispersion for Spanish vowels.

In this investigation, we examine whether the prosodic structure affects vowel quality in L2 French in the speech of learners coming from different L1s (English vs Spanish). We try to determine (a) to what extent the expansion of L2 French vowels depends on the increase of prosodic hierarchy, and (b) whether prosodic strengthening favours the L2 pronunciation accuracy of vowels that are commonly found difficult by learners.

We analyse read speech produced by 30 participants [3, 6]: 10 French native speakers, 10 English learners and 10 Spanish learners (B1 proficiency level of L2 French). Two levels of analysis were retained for examining the effects of prosodic strengthening: vowels produced in (i) IP-final positions, (ii) AP-final edges (obligatory accent) and/or at AP-initial rises (non-obligatory accent on the first syllable of the first content word). We compare vowels produced in these conditions with those produced in simple word-internal non-accented position (WD).

IPs and APs were determined according to a syntax-prosody approach following [10]: the former were associated to coordinated root clauses, extra-sentential elements, main clause edges, etc., whereas the latter consisted in any lexical word and their related grammatical words on its left side. A semi-automatic analysis with Prosogram [7] allowed us to label the vowels produced with melodic movements (glissando threshold: $0.32/T^2$) indicating the realization of the predicted prosodic edges/positions. We measured F1-F2-F3 at the mid point of each vowel for the following set: /i, e, ɛ, a, o, ɔ, u, y, ø, œ/. After filtering aberrant formant detections, we analyzed 12,383 vowels in terms of formant dispersion (convex hull area of F1*F2 and F2*F3 vowel spaces).

The results show that L2 vowels in strong prosodic positions (IP or AP) are more expanded than in weak positions (WD). This observation can be attributed to a positive transfer from L1 in the case of English learners, but probably not for Spanish learners, since [8, 9] claim that pitch accents do not affect vowel quality in this language. Additionally, vowels are clearly more expanded in IP than AP in L1 French. However, the level of prosodic hierarchy in L2 French does not affect vowel dispersion: only marginal differences were found in both groups of learners in these two conditions.

Finally, we performed a Linear Discriminant Analysis (LDA) in order to examine the degree of pronunciation accuracy in L2 [11]. The model was trained on L1 French data, and then used to predict vowel categories on learners' productions. The accuracy of the model's predictions increases by 11% for vowels produced in IP or AP than in WD for both groups of learners. This suggests that prosodic strengthening and L2 pronunciation accuracy could be correlated. We discuss these results in the light of L2 acquisition universals, and the role of prosodic strengthening on the production of new L2 sounds.

References

1. Cho, T. (2005). Prosodic strengthening and featural enhancement: Evidence from acoustic and articulatory realizations of /a, i/ in English. *JASA*, *11*(6), 3867–3878.
2. Cho, T. (2011). The phonetics–prosody interface in laboratory phonology. In: N. C. Kuma, B. Botma, & K. Nasukawa (Eds.), *The continuum companion to phonology*. London/New York: Continuum.
3. Delais-Roussarie, E. & Yoo, H. (2011). Learner corpora and prosody: from the COREIL corpus to principles on data collection and corpus design. *Poznan Studies in Contemporary Linguistics 47* (1), 26-29.
4. Gendrot *et al.*, (2016). Détection automatique d'une hiérarchie prosodique dans un corpus de parole journalistique , *Langue française* 2016/3 (N° 191) 123-149.
5. 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.
6. Herment, S., et al. (2014). AixOx, a multi-layered learners' corpus: automatic annotation. In Díaz Pérez J. & Díaz Negrillo A. (Eds.), *Specialisation and variation in language corpora*, Bern : Peter Lang : 41-76..
7. Mertens, P. (2004). Le prosogramme: une transcription semi-automatique de la prosodie. *Cahiers de l'Institut de Linguistique de Louvain 30*, 1-3, 7-25
8. Nadeu, M. (2014). Stress- and speech rate-induced vowel quality variation in Catalan and Spanish. *Journal of Phonetics*, *46*, 1-22.
9. Ortega-Llebaria, M. & Prieto, P. (2007) Disentangling stress from accent in Spanish: Production patterns of the stress contrast in deaccented syllables. In P. Prieto, J. Mascaró & M.J. Solé (Eds), *Segmental and Prosodic Issues in Romance Phonology*, Amsterdam, Netherlands, Philadelphia PA: John Benjamins, pp. 155–175.
10. Santiago, F. & Delais-Roussarie, E. (2015). The acquisition of Question Intonation by Mexican Spanish Learners of French. In E. Delais-Roussarie, M. Avanzi & S. Herment (eds), *Prosody and Language in Contact: L2 Acquisition, Attrition and Languages in Multilingual Situations*. Heldelberg: Springer, 243–270
11. Wang, H. & van Heuven, V. (2006). Acoustical analysis of English vowels produced by Chinese, Dutch and American Speakers. *Linguistic in the Netherlands*, 237-248.

Durational hand gestures facilitate the learning of L2 vowel length contrasts

Peng Li[1], Florence Baills[1] and Pilar Prieto [1,2]
[1] *Departament de Traducció i Ciències del Llenguatge, Universitat Pompeu Fabra,*
[2]*Institució Catalana de Recerca i Estudis Avançats (ICERA)*

Previous studies have shown the benefits of observing or producing pitch gestures (or hand gestures mimicking pitch movements) to acquire novel L2 tonal and intonational contrasts (see Morett & Chang, 2015; Baills et al., 2018 in press for the learning of lexical tones; see Yuan et al., 2018 for the learning of intonational patterns), as well as the benefits of observing and producing beat rhythmic gestures for improving L2 pronunciation (Kushch et al., submitted and Gluhareva & Prieto, 2017). By contrast, some studies have shown null effects for gestures representing vowel length contrasts in the perceptual discrimination of Japanese vowel length contrasts (e.g., Hirata et al., 2014; Kelly et al., 2017). Yet in practice, some teachers have suggested that visuospatial gestures (encoded horizontally in space rather than vertically) representing duration help learning vowel length contrasts (Roberge et al., 1996; Klein et al., 2010). The main goal of this study is to further investigate the role of durational gestures in the acquisition of L2 vowel length contrasts in Japanese by using a horizontal hand gesture that encodes syllabic differences in length and also by expanding the scope from perceptive measures to potential effects on L2 pronunciation.

In a between-subject experiment with a pre- and post-test design, 50 adult Catalan dominant participants without Japanese knowledge were trained to perceptively identify the Japanese vowel length contrasts and to orally imitate Japanese sentences containing a set of words only contrasting in vowel length. They were assigned to one of the following two training audio-visual conditions, namely (a) Gesture group (i.e., where materials were presented audio-visually with gestures encoding durational information); and (b) No-Gesture group (i.e., where materials were presented audio-visually with no gestures). The experimental procedure is presented in Figure 1. Before and after training, all participants performed both a discrimination task and an imitation task. In the discrimination task, participants were asked to listen to 20 Japanese sentences in which disyllabic target words were embedded, and to decide the length of the second syllable of the target word. In the imitation task, they were asked to repeat 20 Japanese sentences with the target word embedded in the middle position. The speech production of the participants was acoustically analysed for the duration of the target vowel and the mean duration ratio of the long vowel to short vowel was calculated.

The results (see Figure 2) showed that while both groups improved in both tasks after training, the Gesture group yielded larger improvements than the No-Gesture group in perception and production. Two GLMM analyses revealed that there was a significant interaction of Condition*Test for the production, while no such significant interaction was observed for the perception. These results suggest that durational gestures help beginners boost the learning of L2 vowel length contrasts, and that the benefits can be immediately observed in word pronunciation patterns.

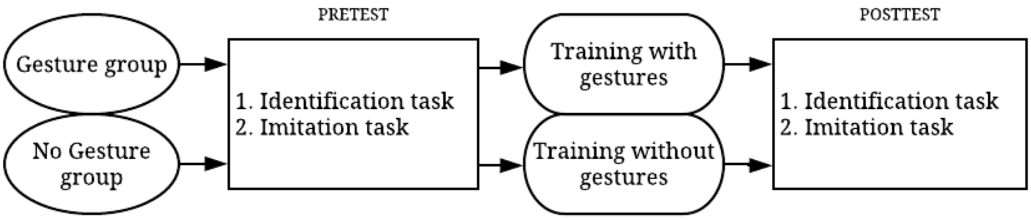


Fig. 1 Experimental procedure

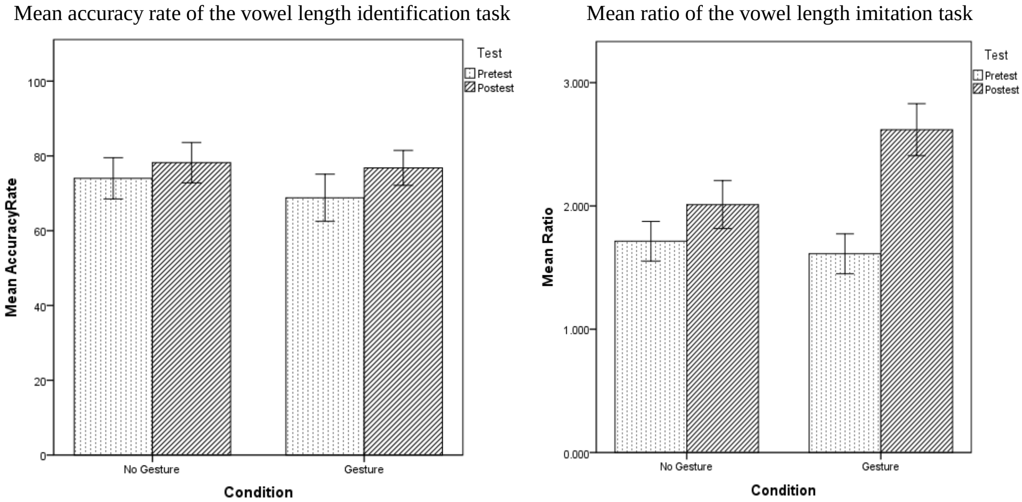


Fig.2 Results of the identification task (left panel) and the imitation task (right panel) across gesture and no-gesture groups in the pre-and post-tests.

REFERENCES

[1] Baills, F., Suárez-González, N., González-Fuente, S., & Prieto, P. (2018, in press). Observing and producing pitch gestures facilitates the acquisition of Mandarin Chinese tones and words. *Studies in Second Language Acquisition*, 1-26.

[2] Gluhareva, D., & Prieto, P. (2017). Training with rhythmic beat gestures benefits L2 pronunciation in discourse-demanding situations. *Language Teaching Research*, *21*(5), 609-631.

[3] Hirata, Y., Kelly, S. D., Huang, J., & Manansala, M. (2014). Effects of hand gestures on auditory learning of second-language vowel length contrasts. *Journal of Speech, Language, and Hearing Research*, *57*(6), 2090-2101.

[4] Kelly, S., Bailey, A., & Hirata, Y. (2017). Metaphoric gestures facilitate perception of intonation more than length in auditory judgments of non-native phonemic contrasts. *Collabra: Psychology*, *3*(1), 7.

[5] Klein, L. (2010). Phonetic correction in class with verbo-tonal method. *Studies in language and literature*, *30*(1), 35-56.

[6] Kushch, O., Gluhareva, D., Borràs-Comes, J., Pérez, C., & Prieto, P. (submitted). The role of beat gesture production in L2 pronunciation training. *Bilingualism: Language and Cognition*.

[7] Morett, L. M., & Chang, L.Y. (2015). Emphasizing sound and meaning: pitch gestures enhance Mandarin lexical tone acquisition. *Language, Cognition and Neuroscience*, *30*(3), 1–7.

[8] Roberge, C., Kimura, M., & Kawaguchi, Y. (1996). *Nihongo no hatsuon shidoo: VT-hoo no riron to jissai* [*Pronunciation training for Japanese: Theory and practice of the VT method*]. Tokyo: Bonjinsha.

[9] Yuan, C., González-Fuente, S., Baills, F., & Prieto, P. (2018, in press.). Observing pitch gestures favours the learning of Spanish intonation by Mandarin speakers. *Studies in Second Language Acquisition*, 1-28.

Keynote speech 3

Turn-taking: Early planning & Late cues

Sara Bögels
Donders Institute for Brain, Cognition, and Behavior, Nijmegen

Conversation is the most important ecological niche for language use. Whereas conversation is indeed the focus of a large part of observational (e.g., CA) and corpus research, experimental studies of language have so far stayed rather far away from it. Apart from complicating generalization of the results to real life language use, this situation moreover has led to the neglect of truly conversational topics, such as turn-taking. Interlocutors in conversation take turns at an amazing speed (with ~200 ms gaps), considering estimates of the much longer time needed for speech planning (> 600 ms). In the last years, I have tried to unravel this psycholinguistic puzzle, using different experimental approaches (e.g., reaction time and EEG studies). In a first series of experiments, I have shown that listeners appear to start planning their response to ongoing questions as soon as they have enough information. However, those listeners that focus on response planning, and can give a quick response, appear to pay less attention to the end of the ongoing turn, whereas late responders show no effect on comprehension. A second series of experiments showed that the fact that interlocutors can respond exactly on time cannot be solely due to long-range (lexical) prediction, but that it also crucially depends on late, mostly prosodic, cues in the ongoing question. In all, these results support a model in which listeners start planning their response early during the previous turn, but rely on late (prosodic) cues to deliver that response on time.

Oral session 8

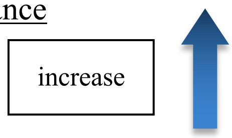
Processing of information

The Incremental Processing of Pitch Accents, Information Status and Focus

Stefan Baumann[1] and Petra B. Schumacher[2]
[1]*IfL Phonetik,* [2]*IDSL 1, University of Cologne, Germany*

The present study investigates the real-time comprehension of items in *First Occurrence Focus* (FOF), *Second Occurrence Focus* (SOF) and *Background* (BG) and their specific prosodic marking in German using event-related potential (ERP) measures. While previous electrophysiological research tested mismatches between prosody and information structure (e.g. [1]), our study assessed contextually licensed, appropriate prosodic realizations.

Crucially, the setup of our study makes it possible to tease apart the independent contributions of focus (defined here morpho-syntactically by a focus particle) on the one hand and information status (here: newness vs. givenness) mediated by context on the other. The three combinations of focus and information status tested also come with distinct prosodic realizations, as previous studies suggest (e.g. [2] show that SOF elements are often marked by phrase accents, i.e. postnuclear prominences expressed by increased duration and intensity but not by tonal movement). Weighting procedures with factors that have a ‘boosting’ or an ‘inhibiting’ influence on an element's prominence have been proposed (e.g. by [3]), assuming the relevance of three distinct levels of prosodic prominence that mirror three distinct levels of information structural weight or importance. (1) translates this relation into a system of binary features.

(1) Information structural importance		Prosodic prominence
FOF (+focus, +new)		Pitch accent (+pitch, +duration)
SOF (+focus, -new)		Phrase accent (-pitch, +duration)
Background (-focus, -new)		No accent (-pitch, -duration)

Forty stimuli per condition were created as part of the answer of a mini dialogue (see examples with the target word *Bier* ‘beer’ in (2)) and read by a trained phonetician. The stimuli (plus 120 filler dialogues) were presented to 21 native speakers of German (17 w, 4 m) in an ERP experiment. A word recognition task served to test participants’ attention to the stimuli. Based on previous ERP research, we predicted increasing processing demands with increasing informational importance and decreasing prosodic prominence [4]. This was widely confirmed.

The manipulation of ***information structure*** revealed increased processing effort over posterior brain regions for FOF items, reflected in a more pronounced negativity between 400 and 650 ms (FOF > SOF/BG; see Fig.1). This supports previous studies (e.g. [4, 5]) and can be attributed to [+new] rather than [+focused] information, i.e. the divide of FOF vs. SOF/BG is between new and given and not between focused and non-focused information.

As to ***prosody***, our results indicate an inverse relation between processing effort and the level of perceived prominence: We found a clear difference over anterior brain regions between the processing of pitch accents (displayed in FOF contexts), which are prosodically prominent due to tonal movement in the vicinity of a stressed syllable, and no pitch accents (comprising phrase accents and deaccentuation in SOF and BG contexts), which lack this tonal movement. This difference was reflected in a biphasic pattern, i.e. a negativity between 250 and 400 ms followed by a positivity between 750 and 950 ms for SOF/BG over FOF. Since increased processing effort is only observed for lack of accents, we assume by implication that the production of a pitch accent, which is more costly for the speaker, reduces the processing costs on the side of the listener. An intermediate status of phrase accents in terms of processing effort and, in turn, prominence perception could not be confirmed.

In conclusion, our data indicate that prosodic and information structural cues influence incremental processing in discrete ways and that pitch accents and newness fulfill specific prominence-lending functions.

(2)

FOF

Context: Was gibt's Neues? ('What's new?')

Target: Karl hat nur **BIER**$_{FOF}$ getrunken. ('Karl only drank BEER.')

SOF

Context: Eva hat nur Bier getrunken. ('Eva only drank beer.')

Target: Sogar THOmas hat nur **BIER**$_{SOF}$ getrunken. ('Even THOmas only drank BEER.')

BG

Context: Wer hat Bier getrunken? ('Who drank beer?')

Target: HANS hat **Bier**$_{BG}$ getrunken. ('HANS drank beer.')

Target words printed in bold face; Capitals indicate fully-fledged (nuclear) pitch accents, small capitals mark phrase accents, and lack of capitalisation indicates complete lack of prominence.

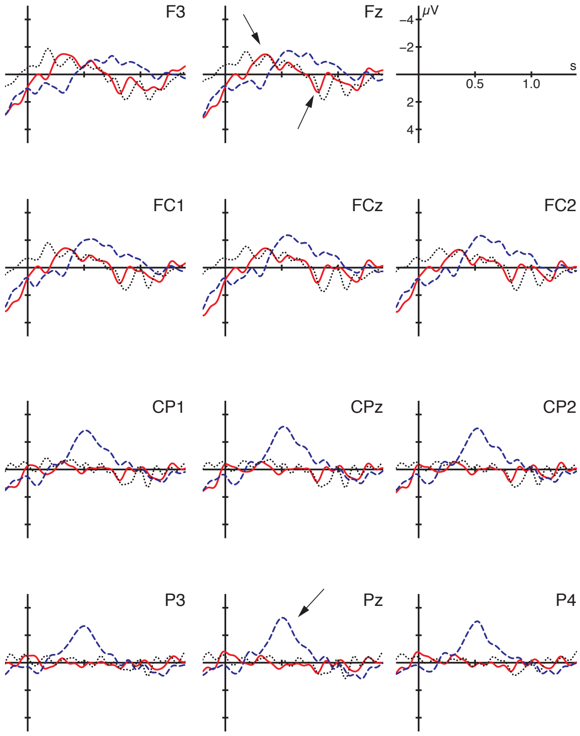


Figure 1. *Grand-average ERPs at selected electrodes for the contrast BG (red solid line) vs. SOF (black dotted line) vs. FOF (blue dashed line), time-locked to the onset of the critical word.*

[1] Toepel, U., Pannekamp, A., & Alter, K. 2007. Catching the news: Processing strategies in listening to dialogs as measured by ERPs. *Behavioral and Brain Functions* 3, ARTN 53.

[2] Beaver, D., Clark, B., Flemming, E., Jaeger, F., & Wolters, M. 2007. When Semantics Meets Phonetics: Acoustical Studies of Second Occurrence Focus. *Language* 83(2): 245-276.

[3 Beaver, D., & Velleman, D. 2011. The Communicative Significance of Primary and Secondary Accents. *Lingua* 121: 1671-1692.

[4] Baumann, S., & Schumacher, P. B. 2012. (De-) accentuation and the processing of information status: evidence from event-related brain potentials. *Language and Speech* 55(3): 361-381.

[5] Burkhardt, P. 2006. Inferential bridging relations reveal distinct neural mechanisms: Evidence from event-related brain potentials. *Brain and Language* 98(2): 159-168.

Pre-activation negativity (PrAN): A neural index of predictive strength of phonological cues

Mikael Roll[1], Pelle Söderström[1,2] and Merle Horne[1]
[1]*Lund University,* [2]*MARCS Institute, Western Sydney University*

Occurring at rates of up to 6–7 syllables per second,[1] speech perception and understanding involve rapid identification of speech sounds and pre-activation of morphemes and words.[2] Using event-related potentials (ERPs) and functional magnetic resonance imaging (fMRI), we have investigated the time-course and neural sources for phonological cues pre-activating word endings[3-6] and syntactic structures[7] as participants hear unfolding words. The ERP results have led us to propose a new brain potential: the 'pre-activation negativity' (PrAN) (Figure 1).[3,4] PrAN is an electrically negative deflection occurring at 136–280 ms following a number of different phonological elements (segmental phonemes,[3] morphologically conditioned word accents,[4-6] and left-edge boundary tones[7]). These phonological elements have in common the fact that they cue relatively few possible word- or sentence-level continuations. Thus, PrAN for word beginnings was found to increase gradually with a decrease in the number of possible word completions[3,4] and an increase in the lexical frequency[3] of those completions. In other words, PrAN became stronger the more certain participants could be about the continuation of a word at a certain point due to a relatively more reduced lexical cohort[8] containing relatively more frequent words. At the clause level, absence of a "left-edge boundary tone" at the beginning of an embedded clause was also shown to produce a greater PrAN as compared to the presence of a boundary tone.[7] The embedded clause context without a left-edge boundary tone led to increased certainty regarding the upcoming syntactic structure of a clause, due to it being associated with fewer syntactic possibilities than if there had been a boundary tone.

For real words, the PrAN for processing of word accents, which are important phonological cues in morphological prediction, correlated with blood-oxygen-level-dependent (BOLD) activity in the left superior temporal gyrus and pars triangularis and orbitalis of the left inferior frontal gyrus (IFG). These are areas of the ventral processing stream, relevant for whole word form access.[9-11] Forcing morphological processing by using pseudowords yielded a more left-frontally distributed PrAN,[12] possibly reflecting involvement of the part of the dorsal stream handling grammatical processing.[13] The PrAN for word beginnings with few possible completions was associated with a BOLD increase in the pars opercularis of the left IFG and angular gyrus of the left parietal lobe. The activated areas are in line with the dorsal brain stream for predictive processing, also engaged during sensorimotor and auditory-motor mapping[9] as well as lexical selection.[14] Left-edge boundary tones cueing syntactic structures activated IFG, pars opercularis, but extended more ventrally than the activation observed for word beginnings.[7]

This presentation will show the time-course and neural underpinnings of the online processing of phonological cues to lexical, morphological, and syntactic pre-activation. Thus, based on ERP-BOLD correlations, we suggest that the PrAN for phonological cues to word forms might be subdivided into an early phase (136 ms),[5,13] involving the ventral processing stream, and a later phase (200 ms), with neural sources along the part of the dorsal stream dealing with sensorimotor and auditory-motor processing.[3,5] The early activation of the ventral stream can be interpreted as representing initial lexical access.[14] The later engagement of the dorsal stream could reflect lexical selection through inhibition of competitors in the activated lexical cohort. The PrAN for phonological cues to grammatical structures seems to have sources in the part of the dorsal stream dedicated to syntactic processing.[15] Finally, PrAN could be thought to be similar to the contingent negative variation (CNV) in that it reflects expectancy for a not yet heard part of a stimulus. However, PrAN's timing is much earlier and rather overlaps the N1 and P2 components, from which PrAN differs in not responding to physical characteristics of phonological cues *per se*, but rather to their predictive potential.

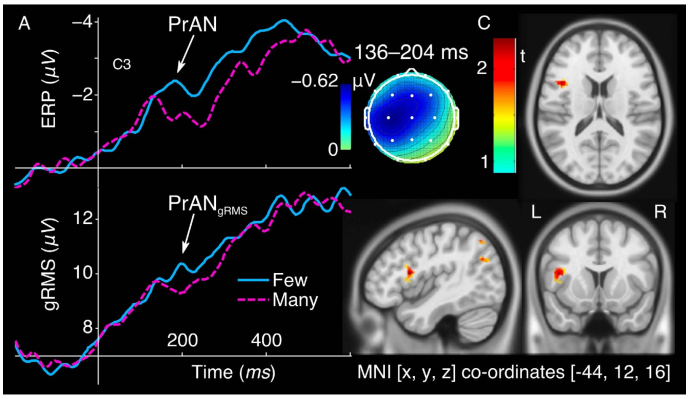


Figure 1. *Pre-activation negativity (PrAN) and possible neural sources at word beginning during online speech processing. PrAN increased with certainty about upcoming word endings, i.e. with fewer possible word completions with higher lexical frequency.*[3]

[1] Levelt, W. J. M. 1989. *Speaking: From intention to articulation*. Cambridge, MA: The MIT Press.

[2] Federmeier, K.D. 2007. Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology* 44(4), 491-505.

[3] Roll, M., Söderström, P., Frid, J., Mannfolk, & Horne, M. 2017. Forehearing words: Pre-activation of word endings at word onset. *Neuroscience Letters* 658, 57-61.

[4] Söderström, P., Horne, M., Frid, J., & Roll, M. 2016. Pre-activation negativity (PrAN) in brain potential activations to unfolding words. *Frontiers in Human Neuroscience* 10, 1-11.

[5] Roll, M., Söderström, P., Mannfolk, P., Shtyrov, Y., Johansson, M., van Westen, D., & Horne, M. 2015. Word tones cueing morphosyntactic structure: Neuroanatomical substrates and activation time-course assessed by EEG and fMRI. *Brain and Language* 150, 14-21.

[6] Roll, M. 2015. A neurolinguistic study of South Swedish word accents: Electrical brain potentials in nouns and verbs. *Nordic Journal of Linguistics* 38(2), 149-162.

[7] Söderström, P., Horne, M., Mannfolk, P., van Westen, D., & Roll, M. 2018. Rapid syntactic pre-activation in Broca's area: Concurrent electrophysiological and haemodynamic recordings. *Brain Research* 1697, 76-82.

[8] Marslen-Wilson, W., & Tyler, L.K. 1980. The temporal structure of language understanding. *Cognition* 8(1), 1-71.

[9] Hickok, G., & Poeppel, D. 2004. Dorsal and ventral streams. *Cognition* 92, 67-99.

[10] DeWitt, I., & Rauschecker, J. P. 2012. Phoneme and word recognition in the auditory ventral stream. *Proceedings of the National Academy of Sciences*. 109, E505-E514.

[11] Schremm, A., Novén, M., Horne, M., Söderström, P., van Westen, D., & Roll, M. 2018. Cortical thickness of planum temporale and pars opercularis in native language tone processing. *Brain and Language, 176*, 42-47.

[12] Söderström, P., Horne, M., Roll, M., 2017. Stem tones pre-activate suffixes in the brain. *Journal of Psycholinguistic Research* 46(2), 271-280.

[13] Söderström, P., Horne, M., Mannfolk, P., van Westen, D., Roll, M., 2017. Tone- grammar association within words: Concurrent ERP and fMRI show rapid neural pre-activation and involvement of left inferior frontal gyrus in pseudoword processing. *Brain and Language* 174, 119-126.

[14] Zhuang, J., Tyler, L.K., Randall, B., Stamatakis, E.A., & Marslen-Wilson, W.D. 2014. Optimally efficient neural systems for processing spoken language. *Cerebral Cortex* 24, 908-918.

[15] Friederici A.D., Bahlmann, J., Heim, S., Schubotz, R.I., & Anwander, A. 2006. The brain differentiates human and non-human grammars: Functional localization and structural connectivity. *Proceedings of the National Academy of Sciences* 103, 2458-2463.

Oral session 9

Sign language

Typological and Historical relations across sign languages
The view from articulatory features

Natasha Abner
University of Michigan

Justine Mertz
Paris Diderot University

Jessica Lettieri
University of Turin

Shi Yu
École Normal Supérieure, Institut Jean-Nicod, EHESS

Carlo Geraci
École Normal Supérieure, Institut Jean-Nicod, CNRS

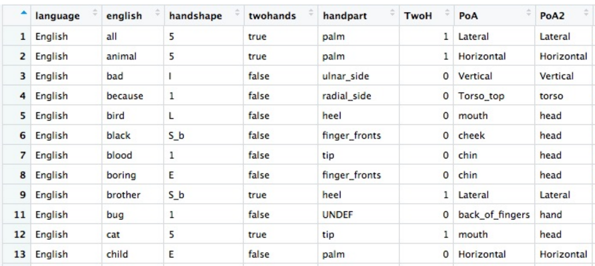
Goals. The aim of this paper is twofold. First, we use articulatory/phonetic features to describe the properties of sign language (SL) lexical items; second, we use these features to classify SLs into macro- and micro-families. The paper provides proof of concept that quantitative methods can be used to probe typological and historical classifications of SLs, along the lines of what has been done in spoken language phylogenetics (Nichols 1992, Dunn et al. 2005) and the genetics of speech communities (Verdu et al. 2017).

Background. SLs are natural languages that are perceived visually (vs. acoustically) and produced gesturally (vs. vocally). The perception-production systems of SL give rise to two macroscopic modality effects. One, SLs allow simultaneous production of a significant amount of contrastive phonemic material (Brentari 1998). Two, SLs display a degree of iconicity at the lexical level, understood here as a conceptually motivated mapping between sign form and sign meaning (Taub 2001 i.a.). Thus, SLs may exhibit a higher degree of cross-linguistic similarity (Wilbur 2010) and signers without a shared language may experience relative ease in converging on a shared communication system (Zeshan 2015). Nevertheless, SLs share many of the structural and functional phenomena of spoken languages. SLs may be classified into language families according to their historical relationships (Anderson & Peterson 1979, Wittman 1991), though additional reliable and verified documentation remains necessary. SLs may also be grouped typologically according to their linguistic properties (Brentari et al. 2015, Zeshan 2006). For example, pairwise comparisons of SLs based on global resemblance of handshape, movement, location and hand-orientation showed that it is possible to detect the degree of similarity/distance between SLs (Woodward 2000; McKee and Kennedy 2000). Here, we assess the efficacy of established statistical models in the typological classification of SLs based on linguistic features. Because typology and history exhibit patterns of convergence and divergence, we also evaluate the typological groupings statistically inferred relative to what is known about historical relatedness among SLs.

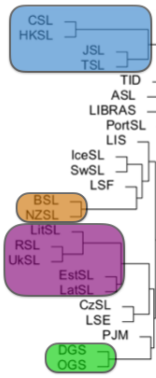
Methodology. To have the same baseline for cross-linguistic comparison, we used Woodward's SL adaptation of the Swadesh list. Following lexicostatistics practice (Rea 1990), this list identifies 100 items that represent some of the core concepts of human life/experience (e.g., *mother,* *live, fire,* etc.). Data from 24 SLs (4 Asian, 7 Eastern European, 9 Western European, plus ASL, LIBRAS, NZSL and TiD) were sourced from an on-line dictionary (www.spreadthesign.com). Articulatory (phonetic) features were manually coded for items on the SL-adapted list for all nine languages (all items were not available for all languages). The set of articulatory features coded (55 handshapes, 36 locations, 11 movements) were modeled after Brentari (1998) but are common across SL phonological models (Sandler and Lillo-Martin 2006, Van der Kooij 2002). The features fall into four major classes (Handshape, Place of Articulation, Hand-orientation and Movement) and distinguish the possible configurations and actions of the hand during sign production. For instance, the feature $[\pm spread]$ distinguishes adjacent vs. non-adjacent fingers in handshapes like and . Coding was done using ad-hoc web-based software for annotation (Author 2, Author 1 and Author 3, 2017). Figure (1) offers an overview of the dataset. Historical information about the SLs in our sample has been retrieved via the Ethnologue of World's languages and the available literature on each language. A graphical representation of the reconstructed historical relations is given in Figure (3).

Analyses. Two types of cluster analyses have been performed on the data: One analysis focuses on major similarities between pairs of signs between pairs of languages. It extends on a large scale the original Woodward's method and it has been successfully applied to spoken languages (Lynch et al. 2002). The results of this method are illustrated in Figure (2). The other analysis focuses on the distribution of articulatory features within the 100 sign sample for each language. It is an adaptation of Dunn et al (2005) method. The adaptation consists in the fact that i) we only considered articulatory features, ii) our data are not in binary form but are counts of each feature. The results of this method are illustrated in Figure (4). Both analyses cluster together i) the group of Asian languages (blue), ii) British and New Zealand SL (orange), iii) Most of the countries of former Soviet Union (purple). The **item analysis** also highlights a relation between Austrian and German SL (green), while the **feature analysis** groups together Austrian, Czech, Lithuanian and Polish SL (green).

1. FIGURE: Overview of the annotated files

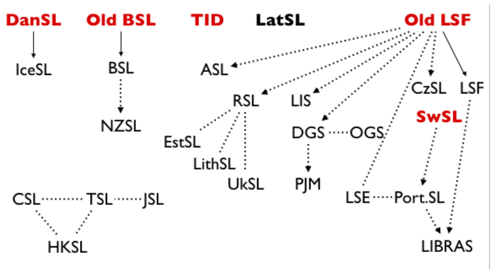


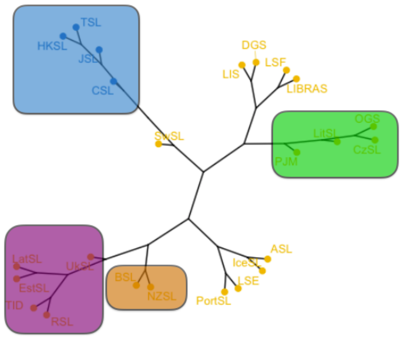
	language	english	handshape	twohands	handpart	TwoH	PoA	PoA2
1	English	all	5	true	palm	1	Lateral	Lateral
2	English	animal	5	true	palm	1	Horizontal	Horizontal
3	English	bad	I	false	ulnar_side	0	Vertical	Vertical
4	English	because	1	false	radial_side	0	Torso_top	torso
5	English	bird	L	false	heel	0	mouth	head
6	English	black	S_b	false	finger_fronts	0	cheek	head
7	English	blood	1	false	tip	0	chin	head
8	English	boring	E	false	finger_fronts	0	chin	head
9	English	brother	S_b	true	heel	1	Lateral	Lateral
11	English	bug	1	false	UNDEF	0	back_of_fingers	hand
12	English	cat	5	true	tip	1	mouth	head
13	English	child	E	false	palm	0	Horizontal	Horizontal



2. FIGURE: Cluster Analysis for items

3. FIGURE: Reconstructed Historical relations





4. FIGURE: Cluster Analysis for features

Discussion. Both analyses partially confirm the reconstructed historical relations. The Asian group can also lead towards an areal interpretation. Under this view, the two cluster analyses witness the fact that Asian SL phonology has features that make it intrinsically different from the other groups. As for the influence of OldLSF, the reason why we do not observe a large cluster with LSF could be due at least to two possible reasons: i) the various languages originated from OldLSF diverged too much over the last two centuries; ii) OldLSF had a similar influence on almost all the remaining SL, which in turns make clusters harder to shape. Evidence of this second hypothesis will be provided during the talk. The green cluster in Figure (4) is somewhat surprising as there is no documentation about it. However, a plausible explanation is available if we look at macrohistory: those countries were formerly part of the Austro-Hungarian Empire. The Empire vanished after World War 1. What our **feature analysis** literally shows are then the fingerprints of human history in the history of these sign languages.

Selected References Anderson, L. & D. Peterson. 1979. A comparison of some American, British, Australian, and Swedish signs: evidence on historical changes in signs and some family relationships of sign languages. • Brentari, D., et al. 2015. Typology in sign languages: Can it be predictive?. • Dunn, M., et al. 2005. Structural phylogenetics and the reconstruction of ancient language history. • Nichols, J., 1992. Linguistic Diversity in Space and Time. • Rea, J. 1990. Lexicostatistics. • Verdu, Paul, et al. 2017. Parallel Trajectories of Genetic and Linguistic Admixture in a Genetically Admixed Creole Population. • Wilbur, R.B. 2010. The semanticsphonology interface. • Wittmann, H. 1991. Classification linguistique des langues signes non vocalement. • Woodward, J. 2000. Sign languages and sign language families in Thailand and Viet Nam. • Zeshan, U. Roots, leaves and branchesThe typology of sign languages. • Zeshan, U. 2015. "Making meaning" - Communication between sign language users without a shared language

Measuring phonological complexity in Sign Languages

[1]Valentina Aristodemo, [1]Chiara Annucci, [1,3]Justine Mertz, [2]Giustolisi Beastrice, [3]Carlo Geraci and [1]Donati Caterina
[1]University of Paris-Diderot, [2]University of Milano-Bicocca, [3]CNRS Institut Jean-Nicod

Goals. We propose a new data-driven measure of sign language (SL) perceptual/articulatory complexity. If satisfactory, the measure will be used to assess in parallel with or in absence of other measures (e.g., frequency and acquisition). We compare the performance of our data-driven measure with a theory-driven measure of complexity based on the SL feature geometry (Brentari, 1998). Our empirical base is composed of four sign languages: French, Italian, Catalan and Spanish (LSF, LIS, LSC and LSE).

Background. There are two main types of approaches to phonological complexity in spoken languages, which we can refer to *theory-driven* and *data-driven*. The *theory-driven approach* is well illustrated by Clements (1985) and Sagey (1986) where counting distinctive features is the relevant measure. *Data-driven approaches* typically rely on pattern/order of acquisition, frequency, speech errors and similar measurable facts to assess phonological complexity. Ideally the two measures should converge.

As for SL *theory-driven approaches* to complexity, Brentari's Prosodic Model (1998) belongs to the former tradition: each phonemic class is assigned a set of features organized in a hierarchical/geometric structure. Each sign can be described in terms of a branching tree. The root corresponds to the prosodic word and branches correspond to the phonemic classes of handshape, location, and movement, each containing its own feature geometry. The richer is the structure (in terms of positively specified features), the higher is the complexity of a sign. While this model provides important crosslinguistic generalizations, its validity beyond ASL cannot be taken for granted. Similar considerations hold for other models (e.g., Sandler & Lillo-Martin 2006 and van Der Kooij 2002).

As for *data-driven approaches*, frequency and acquisition data are available only for a few sign languages (e.g., ASL, BSL, NGT), but they are entirely missing for others and would require long-term efforts to obtain. Diagnostics based on error rates can provide a quick and handy measure for early detection of a number language disorders.

Definitions of complexity. *Our data-driven measure* is based on error rates in naïve non-signers. The rationale is: signs that can be accurately and fluently repeated are treated as simple (see below). *Theory-driven measure*. Adapting Brentari's model, we measured complexity by counting the number of nodes and features necessary to describe it.

Methods. **Data-driven measure**. A repetition task is used to assess sign complexity in non-signers. The procedure is identical for all SLs in the study. We describe here the case of LSF.

Materials. 108 signs in LSF were selected based on criteria such as lack of major iconicity, frequency, lack of regional variation. A Deaf consultant was video-recorded producing the citation forms of the signs.

Participants. 20 hearing non-signers acquainted with the visual culture of France were recruited (divided in 5 age groups: 18-29: 30-40; 40-49; 50-59; 60-70).

Task: Each participant was asked to watch once the video of a sign and (try to) repeat it. Their performance was video-recorded (2160 tokens).

Coding complexity. Two students with a basic competence in LSF coded the data according to 5 measures: fluency, accuracy in handshape, orientation, location and movement. For each measure we assigned 1 if correct, 0 if incorrect. Overall accuracy for a sign is obtained by summing accuracies in each component. The degree of accuracy is directly mapped onto a complexity scale (5 = all correct = least complex, 0 = all incorrect = most complex).

Theory-driven measure: A portion of the entire dataset is used as a pilot study. *Materials*. We annotated 15 items out of 108 used in the data-driven measure. These 15 items received a variable level of complexity in the data-driven measure (5 have a high level of complexity, 5 a low level of complexity and 5 have an intermediate level of complexity).
Coding complexity. The level of complexity depends on the number of nodes and positively specified features in its representation (lower values = less complex, higher values = more complex signs). The total set of nodes and features considered is 116 (handshape=67, location=22, movement =27).
Results. We report here preliminary results form LSF. By the time of the conference, the LSF and LIS data will be available for comparison.
Data-driven method. The overall mean of accuracy is 4.282 (SD=0.82). The most complex sign is HEDGEHOG with an average score of 3.15, while the easiest sign is HAM with a score of 5. Handshape is the class in which most of the errors are observed (45%) followed by movement (39%), orientation (8%) and location (7%). A mixed-model analysis with item and participant as random factors was conducted. A significant main effect of *age* was found (p=0.03). Younger participants are more accurate.
Theory-driven method: our 15 stimuli have an index of complexity that range from 15 to 39. The simplest sign is BAND-AID, while the most complex one is PEN.
Analysis. We observe a correlation between the overall complexity of the theory-driven measure and the overall accuracy/complexity of the data-driven method (r=-0.30). However, it is not significant (p=0.28). We observe a significant correlation between handshape complexity in theory-driven measure and overall accuracy (r=-0.58; p=0.02). The higher the complexity in the handshape is, the lower is the level of accuracy. Other correlations are not significant (movement/location vs. overall accuracy).
Discussion. Preliminary results show that: 1) The two measures converge; 2) Handshape is the class that better correlates with overall accuracy; 3) Still, for some signs we observed considerable divergence. The table illustrates that there are signs that receive a low score in complexity as measured by the theoretical model (i.e., that are predicted to be simple), but still have a poor performance in overall accuracy (e.g., SAUCE), and vice versa (e.g., BONE).

Data-driven scale		Theory-driven scale	
FLOWER	4,9	BAND_AID	16
RED	4,7	SAUCE	21
BAND_AID	4,55	TREE	25
PEAR	4,55	CASTLE	26
BONE	4,5	RED	26
TREE	4,4	PEAR	27
BREAD	4,35	GLASS	28
GLASS	4,2	FLOWER	29
CASTLE	4,1	COMPASS	32
LEAF	3,95	BREAD	32
PEN	3,9	THEATRE	34
FACTORY	3,9	LEAF	35
THEATRE	3,85	BONE	35
SAUCE	3,6	FACTORY	36
COMPASS	3,35	PEN	39

We shall speculate on the source of this divergence. In principle, this could be due to at least one of the following reasons: a) the data-driven measure, being non-linguistic, does not capture some important phonological categorizations; b) the theory-driven measured is not fully equipped to predict complexity in LSF. To address these issues, one could replicate this study in two ways: with signers as participants, and by using pseudo-signs as stimuli. We also expect comparison with the results of the study in LIS to shed light on these issues. Another interesting issue is whether handshape alone is enough to predict complexity. If this is the case, what is the role of place of articulation and movement in determining complexity? One possibility is that location and movement require a fully-fledged phonology in place. In this case, we expect major differences between signers and non-signers.

Selected References: **Brentari**, D. (1998). *A prosodic model of sign language phonology*. MIT Press. **Clements**, G. N. (1985). The geometry of phonological features. *Phonology Yearbook* 2, 225-252. **Sagey**, E. (1986). The representation of features and relations in nonlinear phonology. Doctoral dissertation, MIT, Cambridge, Mass.

Brazilian Sign Language and Articulatory Phonology

Adelaide H.P. Silva and André Nogueira Xavier (*Federal University of Paraná, Brazil*)

It was not until 1960 that American Sign Language (ASL) was demonstrated by William Stokoe to exhibit the same structural principles as spoken languages, and, afterwards, started to be acknowledged as a language on its own [1]. This has attracted other researchers to investigate more deeply ASL and other signed languages across the world. In addition, this has led to the establishment of a new field within Linguistics, where those languages became the object of study.

Because of the difference in their modality of articulation and perception, research on signed languages, especially the pioneering one, took as their main goal to find more and more evidence that these languages are actual languages. They have tried to accomplish this goal, by analyzing signed languages using the same theoretical constructs developed within linguistics for the analysis of spoken languages.

Despite the unquestionable value of all these studies, this approach, not necessary any longer at this point in signed language linguistics history, has overlooked features of these languages which have no (exact) correspondence in spoken languages. Even more damaging than that: By imposing spoken language theoretical constructs on signed language analysis, this approach has disregarded the possibility that these languages follow different principles or even that those analytical categories may be inaccurate even for spoken languages.

This paper aims to take a different approach. On the basis of Articulatory Phonology, grounded in a "general theory of movement", we propose a phonological analysis for Brazilian Sign Language (Libras). More precisely, following [2], we determine the articulatory gestures that constitute Libras lexical items. This task takes into account experimental data obtained by [3]. The author carried out a production experiment that consisted in asking 12 deaf adult signers (six men and six women) from São Paulo city to produce three times in isolation the sign corresponding to a figure they were shown. Signers were videotaped and data were analyzed using [4]. This paper focuses on the gestural representation for one sign selected from the set of data. The selected sign was BEETLE-CAR, and its gestural representation, such as proposed by [5], is illustrated in Figure (1).

In order to achieve our aim, we depart from the assumption that signs are produced in a kind of "tract", which, in our view, are the upper limbs. Therefore, signs produced by the action of both hands are assumed to be produced by two "tracts". Since [2] propose a model for providing a phonological representation of speech sounds, we discuss which articulators would be responsible for the sign production in a signed language such as Libras. Moreover, Articulatory Phonology predicts that articulatory gestures are responsible for implementing discrete tasks involved in a skilled action. Due to the obvious reason that signs differ from speech sounds in relation to the articulators involved in their production, we try to determine the articulators responsible for the production of signs and the discrete tasks they implement.

By doing so, and by assuming that gestures can coordinate in time, we propose a gestural score for the prototypical form of the Libras sign BEETLE-CAR. The coordination of articulatory gestures over time implies that two gestures can superimpose and that this superimposition can be slight or severe, yielding gradient variation in the production of signs.

Besides, different temporal coordination patterns of the gestures involved in the production of a sign, together with the possibility of activation or deactivation of gestures, can explain some of the variation previous research has documented [6], but also other cases of variation that are frequently overlooked and not satisfactorily explained by approaches that, unlike ours, do not analyze signs in terms of units of action.

(1) Gestural score for BEETLE CAR

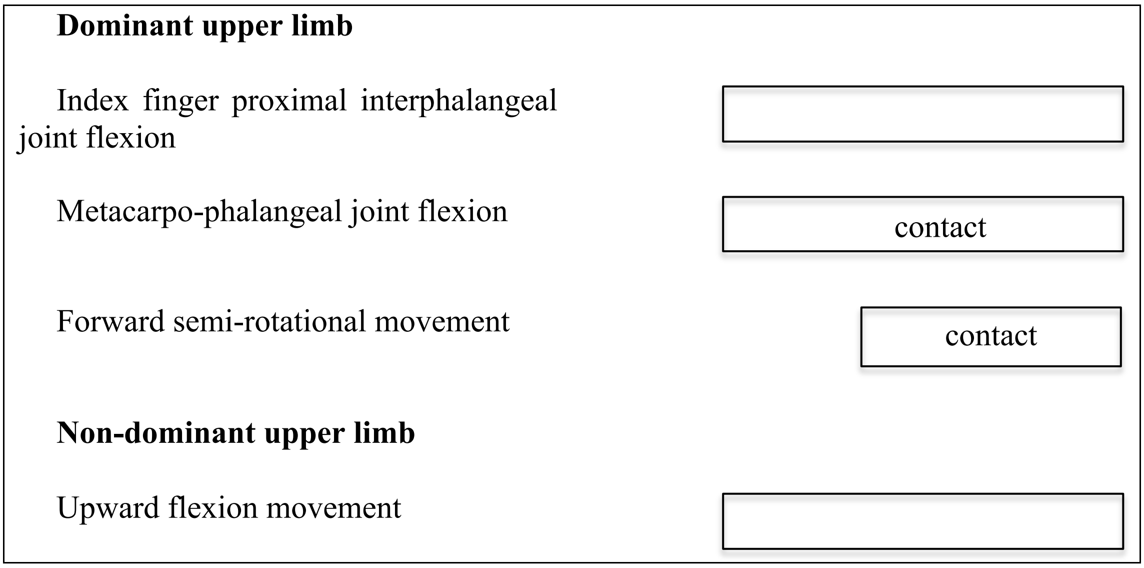


Figure 1. *Gestural score for BEETLE-CAR*.

References

[1] Stokoe, W. 1960. Sign Language Structure: An Outline of the Visual Communication Systems of the American Deaf. Studies in Linguistics: Occasional Papers, 8.

[2] Browman, C., & Goldstein, L. 1992. Articulatory Phonology: an overview. Phonetica, 49, 155-180.

[3] Xavier, A.N. 2014. Uma ou duas? Eis a questão! Um estudo do parâmetro número de mãos na produção de sinais da língua brasileira de sinais (Libras). PhD Dissertation. Unicamp. Unpublished.

[4] ELAN (Version 5.2) [Computer software]. (2018, April 04). Nijmegen: Max Planck Institute for Psycholinguistics. Retrieved from https://tla.mpi.nl/tools/tla-tools/elan/

[5] Silva, A.H.P. & Xavier, A. N. 2018. Libras and Articulatory Phonology. Gradus – Revista Brasileira de Fonologia de Laboratório, 3 (1), 103-124.

[6] Xavier, A.N., & Barbosa, F.V. 2017. Variabilidade e estabilidade na produção de sinais da libras. *Domínios de Lingu@gem*, 11(3), 983-1006.

Poster session 5

Vowel height variation due to prosodic strengthening

Alba Aguete Cajiao[1] and Elisa Fernández Rei[1]

[1]*Instituto da Lingua Galega / Universidade de Santiago de Compostela*

Vowel reduction is one of the most noticeable characteristics of the unstressed vowel system of Galician [9]. This reduction yields the neutralization of mid-high and mid-low vowels in both front and back mid vowels, /e/-/ɛ/ and /o/-/ɔ/. However, some authors [7, 8] pointed out that there is an exception to this reduction process that leads to the lowering of word initial mid vowels of neologisms (i.e. [ɛˈlɛktɾikʊ] *electric*, [ɔˈβesʊ] *overweight*) which contrasts with the traditional mid-high vowels in such position (i.e. [o̞ˈβeʎa̠] *sheep*, [e̞ˈsowtɾʊs] *those others*) [9]. Moreover, it seems that there is some variation on the degree of lowering, and speakers reveal having a certain degree of uncertainty in their production of unstressed word-initial vowels. Despite the effort made by some authors to describe this exception the reason for the lowering of the initial vowels remains an unsolved question.

Prosodic constituents are domains of application for some phonological rules [3]. Furthermore, some authors have investigated variation within the edges of prosodic domains, and they found domain-initial strengthening for consonants [4] and vowels [5] in different languages. This leads us to consider that the position of the segment within the prosodic domain can affect vowel articulation in Galician, causing the aforementioned lowering.

The aim of this research is to examine domain-initial strengthening for word-initial vowels at the beginning of different prosodic domains, namely: prosodic word, phonological phrase, intonational phrase and phonological utterance. The hypothesis we propose is that initial boundary position in the prosodic domain will trigger the strengthening of the initial vowel.

In order to verify our hypothesis, a production experiment with eleven female native speakers of Galician has been carried out. Speakers read a series of statements where the tested vowel ([e], [ɛ], [o] and [ɔ]) appeared in a [V'CVCV] structure at the beginning of the aforementioned prosodic domains. Vowel segments were acoustically analyzed, considering F1 and F2 as they are the primary cues concerning the vowel height contrast. Data was then normalized using the Lobanov method [1]. A linear mixed effect analysis (LMM) of the relationship between the formant values –F1 and F2– and prosodic domain has been performed in R with *lme4* [2] for each vowel, with an intercept for participant as a random factor. A intercept-only models were carried out since it takes account for the residual error related to the subject without overfitting the model [6].

Results show a significant main effect of *domain* in both the F1 and the F2 for the four vowels (Figures 1 and 2). Pairwise comparisons reveal that the higher the prosodic domain is in the hierarchy, the higher the F1, meaning a gradual lowering. Moreover, for the front vowels, the higher the prosodic domain, the higher the F2, whereas for the back vowels, the higher the prosodic domain, the lower the F2, meaning more extreme realizations within the higher domains (Figure 2). In spite of those trends, the prosodic word has not behaved consistently.

Those results have several implications on the phonological processes related to the unstressed vowel system of Galician. On the one hand, they do not support the existence of seven phonological unstressed initial vowels, as various authors stated. However, it seems that there is a neutralization towards the mid-high vowel, at least for the mid back vowels. Furthermore, results show that there is phonetic variation within unstressed mid front and mid back vowels which is triggered by the prosodic structure.

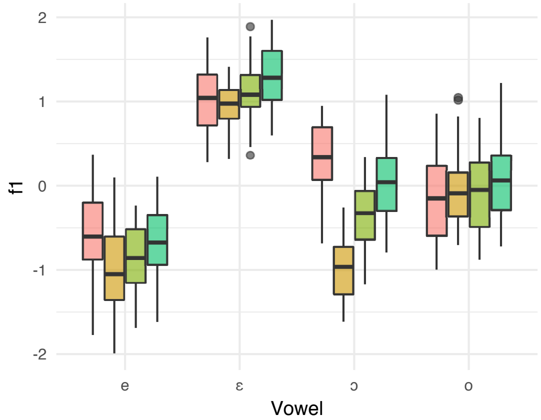


Figure 1a. Boxplots showing F1 values for each vowel within the initial boundary of four prosodic domains tested.

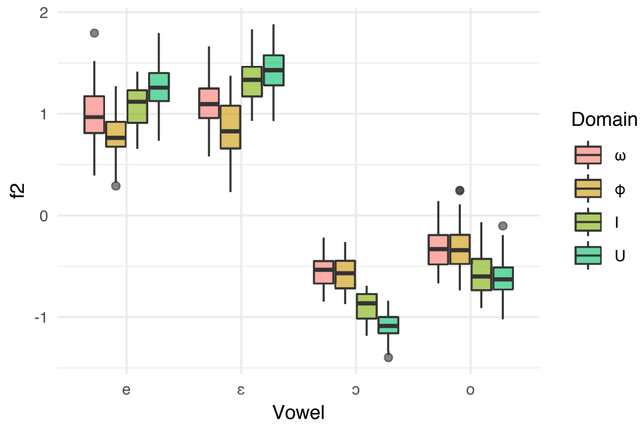


Figure 1b. Boxplots showing F2 values for each vowel within the initial boundary of the four prosodic domains tested.

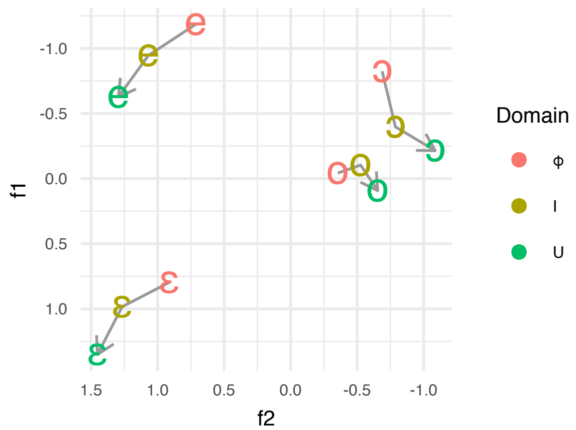


Figure 2. Average normalized F1 and F2 frequencies for each vowel and domain (excluding the prosodic word).

[1] Adank, P., Smits, R. & van Hout, R. 2004. A comparison of vowel normalization procedures for language variation research, *J. Acoust. Soc. Am*. 116, 3099–3107.

[2] Bates, D., Maechler, M., Bolker, B. & Walker, S. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67(1), 1-48. doi:10.18637/jss. v067.i01.

[3] Fernandez Rei, E. 2002. *Regras fonolóxicas posléxicas e regras precompiladas de alomorfia sintagmática, dominios prosódicos*. PhD dissertation. Departamento de Filoloxía Galega: Universidade de Santiago de Compostela.

[4] Fougeron, C. & Keating, P. 1997. Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America* 101 (6), 3728-3740.

[5] Georgeton, L. & Fougeron, C. 2014. Domain-initial strengthening on French vowels and phonological contrasts: Evidence from lip and spectral variation. *Journal of Phonetics* 44, 83-95.

[6] Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H. & Bates, D. 2017. Balancing Type I Error and Power in Linear Mixed Models. *Journal of Memory and Language* 94, 305-315.

[7] Regueira Fernández, X. L. 2009. Cambios fonéticos e fonolóxicos no galego contemporáneo. *Estudos de Lingüística Galega* 1, 147-167.

[8] Santamarina, A. 1972. El verbo gallego: Estudio basado en el habla del valle del Suarna. *Verba*, Anexo 4. Santiago de Compostela: Universidade de Santiago de Compostela.

[9] Veiga, A. 1976. *Fonología gallega*. Valencia: Bello.

Experience with an uptalk variety and perception of high rising terminal contours

Yuki Asano[1,2], Anne Cutler[3], Ann-Kathrin Grohe[4], Andrea Weber[1]
[1]*University of Tübingen,* [2]*Kobe University,* [3]*Western Sydney University,* [4]*Nuance Communication*

Prosodic differences among the varieties of a single language have only recently attracted research attention, and in particular little is yet known about their implications for perception, especially by users or learners of another variety. English varieties, for example, differ in the use and realization of uptalk (high rising terminal contours, or HRTs, in statements [1]). These have been reported more frequently for Australian than for other English varieties [1-3], and Australian English (AusEng) L1 listeners tend to interpret rapid rises from a low onset as statements [4]. Studies of the use of uptalk in conversation [5-10] have examined the roles of speaker ethnicity, age, and sex, and listener sex, but no study to date has addressed effects of variety-specific experience. Here we test perception of AusEng HRTs by L1 listeners from the same versus from a different variety, and by L2 learners with and without experience of AusEng. Note that HRTs do also occur in other languages, albeit less often [1,11].

Participants were 24 AusEng L1 users, 24 American English (AmEng) L1 users, 16 highly proficient German learners of English with at least eight month residence in Australia, and 24 highly proficient German learners of English without such experience. A female AusEng speaker recorded 19 sentences, all having the same elliptical syntactic structure ‘Got a NOUN’ (NOUN = e.g. *flower*, *piano*) that could be interpreted either as a statement “I’ve got a NOUN.” or as a question “Have you got a NOUN?”; a different noun occurred in each sentence. The sentences were manipulated along three binary phonetic variables: low vs. high rise onset, low vs. high rise offset and early vs. late rise onset, giving 152 trials (19 nouns x 8 contours) in total. These were presented auditorily in random order to the participants, who made a categorical judgment for each sentence as to whether it was a statement or a question.

AusEng users made 36% statement responses. With this as the intercept in a mixed effects analysis, German learners without AusEng experience produced the highest proportion of statement responses (50%, $ß = -0.68$, $SE = 0.19$, $z = -3.6$, $p < .001$), followed by AmEng listeners (45%, $ß = -0.46$, $SE = 0.13$, $z = -2.7$, $p < .001$) and then German learners with AusEng experience (41%, $ß = -0.35$, $SE = 0.22$, $z = -1.62$, $p = .1$). Thus the experienced German listeners did not differ in overall proportion of statement responses from the AusEng L1 group, while the two other groups did. Nevertheless, the three non-Aus groups neither differed from one another ($ß = 0.33$, $SE = 0.23$, $z = 1.5$, $p = .1$ between the German groups, $ß = 0.22$, $SE = 0.18$, $z = 1.2$, $p = .2$ and $ß = 0.11$, $SE = 0.21$, $z = 0.5$, $p = .6$ between AmEng and Germans without and with AusEng experience respectively). Further, the non-AusEng groups’ mean responses did not differ significantly from chance level, see Fig. 1. Analyses of phonetic factors revealed that only AusEng listeners gave more statement responses in the low-rise onset condition compared to the high one, see Fig. 2. This pattern corroborates [5] in that low-rise onsets most strongly determined the perception of HRTs. AusEng L1 differed from the three non-AusEng groups in this: vs. Germans with AusEng experience, $ß = 0.36$, $SE = 0.11$, $z = 3.2$, $p < .01$; vs. AmEng, $ß = 0.58$, $SE = 0.10$, $z = 6.2$, $p < .01$; vs. Germans with no AusEng experience, $ß = 0.29$, $SE = 0.10$, $z = 2.9$, $p < .01$.

Thus only the AusEng listeners were sensitive to the different phonetic forms presented in this study. Listeners with no experience of AusEng, whether listening in their L1 or L2, gave random chance-level responses. A small but statistically weak tendency was observed for L2 listeners with AusEng listening experience to converge towards the L1 pattern in that their responses were somewhat less random overall. This suggests that much more experience (than our group's minimum of eight months) with the variety is needed before the precise phonetic realization of uptalk can be correctly identified, and even the relative (in)frequency of uptalk in daily listening can be accurately apprehended.

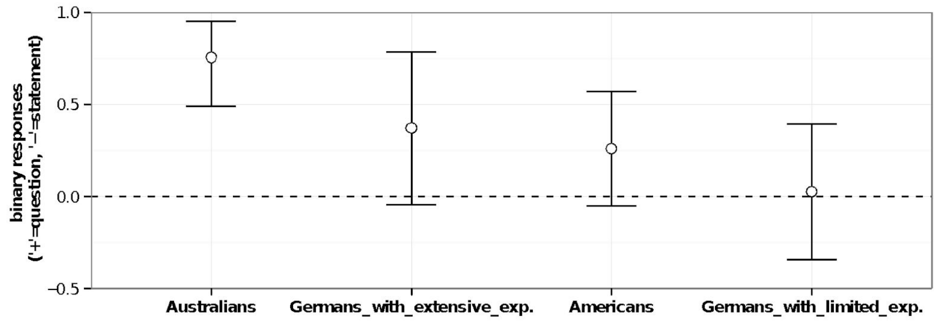


Figure 1: Mean binary responses per group (error bars: 95% CI). The higher the y-axis value is, the higher is the proportion of 'question'-responses.

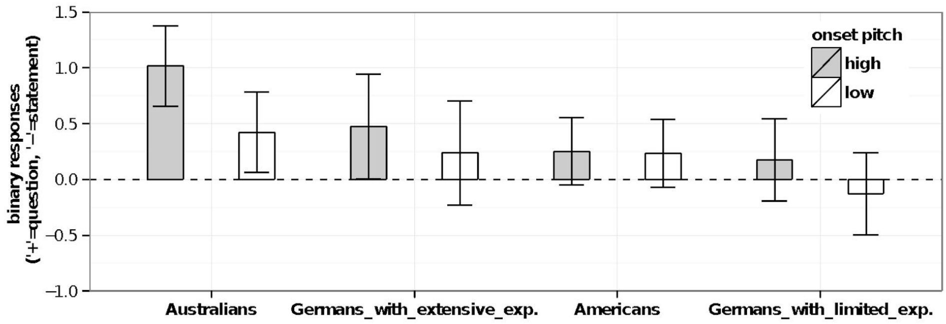


Figure 2: High vs. low onset pitch: mean binary responses per group (error bars: 95% CI).

References

[1] Warren, P. 2016. *Uptalk*. Cambridge University Press, Cambridge (UK).
[2] Fletcher, J., & Harrington, J. 2001. High-rising terminals and fall-rise tunes in Australian English. *Phonetica*, 58, 215-229.
[3] Fletcher, J., & Loakes, D. 2006. Patterns of rising and falling in Australian English. *Proceedings of the Eleventh Australasian Conference on Speech Science and Technology*. (Canberra, Australia), 42-47.
[4] Fletcher, J., & Loakes, D. 2010. *Interpreting rising intonation in Australian English*. *Proceedings of Speech Prosody 2010* (Chicago IL, USA).
[5] Di Gicacchino, M., & Crook Jessop, L. 2010. Uptalk: towards a quantitative analysis. *Toronto Working Papers in Linguistics*, 33, 1-15.
[6] Shokeir, V. 2007. *Uptalk in Southern Ontario English*. Paper presented at NWAV36, University of Pennsylvania.
[7] Szakay, A. 2007. *Identifying Maori English and Pakeha English from suprasegmental cues: A study based on speech resynthesis*. MA thesis, University of Canterbury, NZ.
[8] Szakay, A. 2008. Suprasegmentals and ethnic stereotypes in New Zealand. In P. Warren (ed.), *Laboratory Phonology 11 Book of Abstracts*. Wellington, NZ; pp. 135-136.
[9] Szakay, A. 2012. Voice quality as a marker of ethnicity in New Zealand: from acoustics to perception. *Journal of Sociolinguistics*, 16(3), 382-397.
[10] Ulbrich, C. 2008. Acquisition of regional pitch patterns in L2. *Proceedings of Speech Prosody 2008*, (Campinas, Brazil), 575-578.
[11] Flaig, S., & Zerbian, S. 2016. Rising intonation in statements of German speakers of English as a Foreign Language. Paper presented at *Tone and Intonation in Europe 2016*.

The prosodic word as the domain of French accentuation - Empirical evidence

Corine Astésano
Octogone-Lordat, Université de Toulouse, UT2J, Toulouse, France

French prosodic characteristics are particularly challenging for phonological theories. Our contribution aims at discussing some core issues in French prosodic phonology in the light of a series of results from production, perception and neuroimaging experiments.

Current models of French accentuation unanimously consider French accentuation as being post-lexical, with a primary final accent (FA) and a secondary initial accent (IA) marking the *ap* level. There is, however, no clear consensus as to the respective status of both FA and IA. Whereas only Post [1] considers IA as a pitch accent like FA, most authors describe IA (LHi) as a 'loose boundary marker' because its peak can be aligned with up to the third syllable of a word in a long *ap* [2; 3]. Only the L tone of LHi is consistently aligned with the beginning of the lexical word. IA is also secondary insofar as it is said to yield to FA in case of tonal crowding on shorter constituents, and essentially has a rhythmic function [2; 4]. The status of FA is not entirely clarified either: although it is clearly a *pitch accent* at the *ap* level for most current models, some descriptions suggest that it may survive at higher levels of prosodic constituency [1: LH*-H%; 4], while the most widespread view suggests that it looses its metric quality at the *IP* level in favour of the sole boundary tone H% [5; 2]. This latter proposition stems from a phonological phenomenon quite specific to French, *i.e.* the syncretism between accent (LH*) and intonation contours, which blurs the clear acoustic realization of FA at the surface level. Because stress is also not lexically distinctive, it has lead to the qualification of French as a 'boundary language' [6; 7] at the post-lexical level, or even as a 'language without accent' [5].

The main points we wish to make here are 1) that the essentially demarcative function of French accentuation does not preclude the *metrical* reality of both initial and final accents; and 2) that both accents surface in turn or concomitantly to mark prosodic constituency as early as the lexical/prosodic word level, just below the *ap* level. Because the powerful tonal formalism of *AM*, allowing the distinction between pitch accents (*) and boundary tones (%), is challenged by the peculiarities of French prosodic phonology, we argue that attempts to disentangle stress from boundary phenomena in French should also incorporate rhythmic and durational phenomena [8; 9]. We also argue that the transcription of tonal phenomena needs to be complemented with perceptual analysis by naïve listeners. Perception is viewed here as an interface between acoustic-phonetic cues and phonology [10]. We will present a series of experimental results accounting for the respective role of FA and IA in the marking of prosodic constituency. The focus will particularly be on perceptual results, which help shed light on several phonological issues of French prosody (Figure 1), showing that: 1) IA consistently marks the prosodic structure and more readily so than FA (as shown on production data: [11]); 2) IA's saliency is consistently perceived on the first syllable of the lexical word, irrespective of its peak's (Hi) alignment in the unit, thus further reinforcing the metric interpretation of IA [12]; 3) FA is perceived as metrically more salient than unstressed syllables even at higher levels of the prosodic hierarchy (typically, *IP*) and independently from boundary tones [12]; 4) FA metrical weight 'survives' above the *ap* level, whatever its tonal realization (rising, falling or downstepped) [12; 13]; 5) Downstepped realization of FA (!H) is accounted for at different levels of prosodic hierarchy: *within the ap* (at the *pw* level), it is perceived as metrically stronger than unstressed syllables, and when marking the *IP* level, !H modulates boundary strength perception [13]; 6) Finally, neuroimaging results show that French listeners process stress, indicating both the automaticity of stress extraction and an expectation for words to be stressed in the pre-lexical stage of speech processing [14].

Altogether, these results not only question the notion of French listeners' stress deafness [15], but also advocate for the metric quality of both IA and FA at all levels of prosodic hierarchy. More importantly, they indicate that French prosodic phonology needs to integrate the level of the prosodic word [16] to encompass the whole extent of accentuation rules.

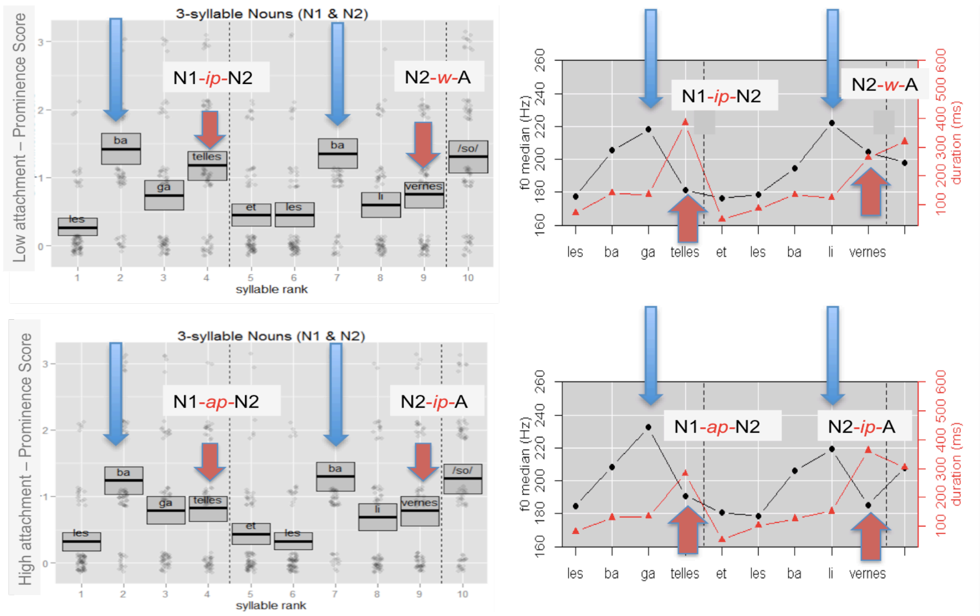


Figure 1. *Perception (left) and production (right : f0 in black and duration in red) of the same sentence in two syntactic conditions, yielding different levels of prosodic constituency. Blue arrows indicate IA and red arrows indicate FA.*

[1] Post, B. 2000. *Tonal and Phrasal Structures in French Intonation*. Thesus, The Hague.
[2] Jun, S.-A. ; Fougeron, C. 2000. A phonological model of French intonation. In: Botinis, A. (Ed.). Intonation: Analysis, Modelling and Technology. Kluwer, Boston: 209–242.
[3] Welby, P. 2003. *The slaying of Lady Mondegreen, being a study of French tonal association and alignment and their role in speech segmentation*. PhD dissertation, OSU.
[4] Di Cristo, A. 2000. Vers une modélisation de l'accentuation en français. Deuxième partie : le modèle. *Journal of French Language Studies*, 10, 27-44.
[5] Rossi, M. 1980. Le français, langue sans accent ? In I. Fónagy & P. Léon (eds.) : *L'accent en français contemporain (Studia Phonetica)*, 15, 13-51.
[6] Vaissière, J. 1990. Rhythm, accentuation and final lengthening in French. In : J. Sundberg, L. Nord & R. Carlson (Eds.). *Music, language, speech and brain,* 108-120.
[7] Beckman, M.E. 1992. Evidence for Speech Rhythms across Languages. In Tohkura; Vatikiotis-Bateson; Sagisaka (eds.) : *Speech Perception, Production and Linguistic Structure*. Tokyo: 457-463.
[8] Dilley, L., Breen, M., Gibson, E. *et al,* 2006. A comparison of inter-coder reliability for two systems of prosodic transcriptions: RaP and ToBI. *INTERSPEECH.*
[9] Astésano, C., & Bertrand, R. 2016. Accentuation et niveaux de constituance en français: enjeux phonologiques et psycholinguistiques. *Langue Française*, 191 (3), 11-30.
[10] Cole, J., Mo, Y. & Hasegawa-Johnson, M. 2010. Signal-based and expectation-based factors in the perception of prosodic prominence. *Laboratory Phonology*, 1(1): 425–452.
[11] Astésano, C.; Bard, E.; Turk, A. 2007. Structural influences on Initial Accent placement in French. *Language and Speech*, 50 (3), 423-446.
[12] Astésano, C.; Bertrand, R.; Espesser, R.; Nguyen, N. 2012. Perception des frontières et des proéminences en français. *JEP-TALN-RECITAL 2012*, Grenoble, 4-8 juin: 353-360.
[13] Mendez, R. ; Astésano, C. (2017) Perception of the Downstepped Final Accent in French. *Phonetics and Phonology in Europe, PaPE 2017,* Köln : 102-103.
[14] te Rietmolen, N., El Yagoubi, R. Ghio, A., & Astésano, C. 2017. The phonological status of the French initial accent and its role in semantic processing: an Event-Related Potentials study, *Interspeech 2017,* Stockholm : 2436-2440.
[15] Dupoux, E., Pallier, C., Sebastian, N., & Mehler, J. (1997). A destressing "deafness" in French? *Journal of Memory and Language*, 36(3), 406-421.
[16] Selkirk, E. (1996). The prosodic structure of function words. In: K. Demuth and J. Morgan: *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*: 187-213.

Variability in the dynamic of nasal vowels in European Portuguese

Conceição Cunha[1], Samuel Silva[2], Arun Joseph[3] António Teixeira[2], Catarina Oliveira[2], Paula Martins[2],Jens Frahm[3])

[1]*IPS, LMU Munich, Germany* [2] *Aveiro University, Portugal* [3]*Max Planck Institute for biophysical Chemistry, Germany*

The main aim of this paper is to describe the dynamic of nasal vowels in two segmental contexts and two prosodic conditions. Further issues are to relate different temporal dynamics to the size of the phonological inventories (see Oliveira, 2009, Martins, 2014) and to better understand the relationship between acoustic and articulatory data. The main hypothesis to be investigated is whether the coordination of oral and nasal gestures is delayed in Portuguese (Oliveira, 2009, Meireles et al, 2015).

The analysed corpus consists of minimal pairs containing all stressed oral [i, e, ɛ, a, ɔ, o, u] and nasal vowels [ɐ̃, ẽ, ĩ, õ, ũ] in one and two syllable words. All words were randomized and repeated in two prosodic conditions embedded in one of three carrier sentences alternating the verb and adverb as follows (*Diga*| *ouvi*| *leio* 'Say| I heard| I read' and *baixinho*|*depois 'gentle, after'*) as in '*Diga ponte, diga ponte baixinho*' ('Say bridge, Say bridge gently'). So far, this corpus has been recorded from fifteen native speakers of European Portuguese (8m, 7f) of EP and the work presented here considers the articulatory data already processed for two speakers.

Real-time magnetic resonance imaging (RT-MRI) recordings were conducted using a 3 Tesla Siemens Prisma Fit MRI System equipped with a 64-channel head coil (Niebergall et al. 2013, Frahm et al. 2014). MRI acquisitions involved a low-flip angle gradient-echo sequence with radial encodings and a high degree of data undersampling. Speech sound was synchronously recorded by means of an optical microphone (Dual Channel-FOMRI, Optoacoustics). Additional data (20 speakers) was recorded without scan noise. The RT-MRI data was processed using the semi-automatic methods proposed by Silva and Teixeira (2015) resulting in sequences of vocal tract contours (see Fig 1.). The analysis framework (Silva and Teixeira, 2016) proposes the comparison of different vocal tract configurations by dividing them in multiple regions with articulatory relevance (e.g., velum, tongue tip, tongue back, tongue dorsum) and applying measures that provide a normalized quantitative assessment of regional differences. Static analysis, as in Fig 2, is introduced for a clarification of the framework capabilities. Here, we can see differences on the vocal tract (VT) between the production of [a] and [u]. The diagram answer the question: "What happens to the vocal tract when I move from an [a] to an [u]?". The major differences between the VT configurations of both vowels were quantified and are visually presented, in the right panel, and include the movement of the lips, with a reduction of the lip aperture (LA) for the rounding of the [u]; the tongue dorsum (TD) moves up; the tongue tip (TT) position changes slightly back and down; and the tongue back (TB), as a result of a change in tongue height, moves further from the pharynx wall. As expected, no relevant differences in the velar region are observed.

For a first overall dynamic analysis of nasal vowels, to get a general insight on the major aspects of their production, Fig. 3 shows, as an example, the evolution of the vocal tract configuration along the production of [ẽ], i.e., what changes happen in the VT for the time interval annotated as an [ẽ]. The production of [ẽ] is essentially characterized by velum movement, with a notable change around 33% of the vowel duration, and some variation for the tongue dorsum and tongue tip, observed for the end of the vowel, and a possible consequence of coarticulation effects. Acoustically, the entire first two formants (F1 and F2) were compared for each nasal vowels and oral counterpart in order to identify major modifications in formants trajectories. Although the analysis is still on going, preliminary results confirm that the velum closure starts during the first half of the acoustic vowel and it is strongly influenced by vowel quality and prosodic context.

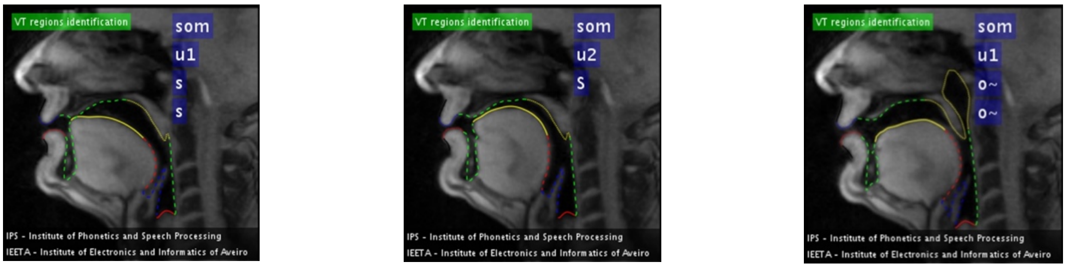


Figure 1. *Vocal tract contours of the alveolar [s] and postalveolar [ʃ] sibilants and the nasal vowel [õ] in the word som 'sound' for comparison reasons.*

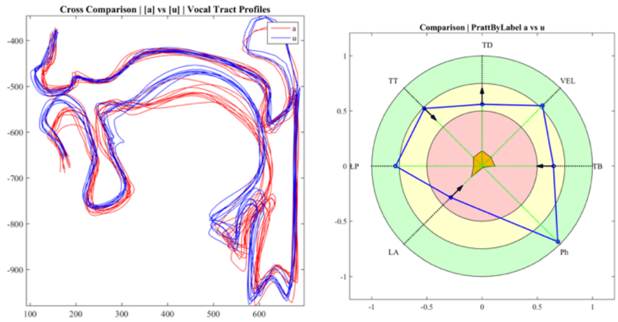


Figure 2. *Crosscomparison of the vowels [a] an [u] : Vocal tract contours in the left panel and quantified normalized regional differences between the two vowels, in the right panel. On the right, a dot in the yellow or red circular coronas signals notable differences for the correspondig region, with a small arrow hinting the direction of the displacement for that articulator.*

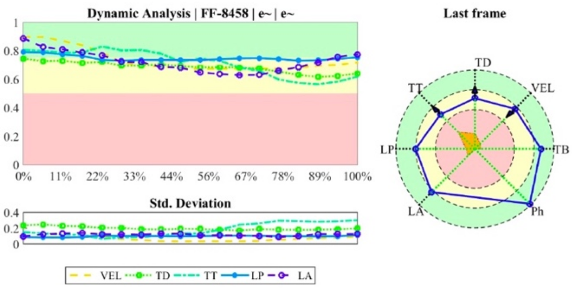


Figure 3. *Movement to the different articulators along the production of [ẽ] gathering data (i.e., occurrences) for multiple contexts. The variability observed, at the end of the vowel, is probably caused by coarticulatory effects*

Selected references

Frahm, J., Schätz, M. Untenberger, M., Zhang, S., Voit, D., Merboldt, K. D. Sohns, J. M., Lotz, J. and Uecker, M. 2014. On the temporal fidelity of nonlinear inverse reconstructions for real-time MRI–the motion challenge. *The Open Medical Imaging Journal*, 8, 1–7.

Niebergall, A., Zhang, S., Kunay, E., Keydana, G., Job, M., Uecker, M. and Frahm, J. (2013). Real-Time MRI of speaking at a resolution of 33 ms: Undersampled radial FLASH with nonlinear inverse reconstruction. *Magnetic Resonance in Medicine*, 69, 477-485

Silva, S. and Teixeira, A. 2015. Unsupervised segmentation of the vocal tract from real-time MRI sequences, *Computer Speech and Language*, 33/ 1, 25–46.

Silva, S. and Teixeira, A., 2016, Quantitative Systematic Analysis of Vocal Tract Data, *Computer Speech and Language*, 36/C, 307-329

Rise shape dynamics is sufficient to distinguish question and continuation rises in French

Mariapaola D'Imperio[1,2] & Lydia Dorokhova[2]

[1]*Rutgers University, New Brunswick (NJ), USA*
[2]*Aix Marseille University, CNRS, LPL UMR 7309, Aix en Provence, France*

Two rising French contours, both labelled as H*H%, are used respectively as a continuation rise or a yes/no question terminal in French [1, 2, 3]. The two contours also differ in terms of duration, with continuation rises showing longer pre-boundary lengthening [4, 5] while H% height has not been found to be consistently higher in questions [3]. More crucially, since Delattre's [6] initial observations, the two rises have been suggested to differ in shape, with continuation contours being more concave than questions', though this has not been confirmed by subsequent perception studies [7] or else the opposite effect has been found [8].

The AM model of intonation stipulates that the shape of the interpolation as well as the slope between two tonal targets is not relevant for descriptive purposes [9] (see [10] for a review). Dynamic properties of the F0 contour, such as peak shape [11, 12] and slope [11] [13] have on the other hand been claimed to be perceptually relevant for intonation category assignment. For instance, the Tonal Center of Gravity [14] hypothesis is based on the idea that perceived properties of intonation contours might be influenced by auditory integration of the dynamic properties of the contour, such as the area covered by a high or low F0 region within the pitch accented syllable.

The hypothesis tested in this study is that a perceptual difference between the two rising contours of French could stem from the particular dynamic characteristics of the rise, notably from the shape of the interpolation between the L and the H tonal targets, once duration and tonal target scaling are kept equal. The same segmental content was hence used to create two intonational bases, produced by a female French speaker as either a declarative question or as a non-final, continuation contour ("La beauté des canaux?/," – *The beauty of the canals?/,*). Segmental durations as well as tonal target scaling and alignment were rendered ambiguous between the two bases in stylized versions through PSOLA resynthesis in Praat [15]. Then, the form of the interpolation between the L (199 Hz) and the H (379 Hz) targets was modified by adding a midpoint whose scaling was altered in 17 steps of 10 Hz, from 209 Hz to 369 Hz, creating several rise shapes from the most convex to the most concave (see Fig.1).

Thirty native French speakers (6 males) participated in a two forced-choice test (continuation vs question response) with the help of the Perceval software [16], and with reaction times being recorded. We predicted that question scores would differ as a result of the Step (17 steps) manipulation as well as a function of the Base (question vs. continuation). Fig. 2 shows that at step 10, continuation stimuli started to be perceived as questions above 50% chance, while for questions this was true already after step 5. Mixed-effects regression models (with Participants treated as random factors, with random slopes) were applied to test the significance of the results. The model retained both Step and Base as significant fixed factors, while their interaction was not significant. Note that percentage of question responses increased with Step in both continuation-base (β =0.18 SE β =0.02, $p<0.0001$) and question-base (β =0.16 SE β =0.02, $p<0.0001$) stimuli, though continuation base stimuli globally received less question responses.

Our perception results reveal for the first time that contour shape alone can play a significant role in the identification of intonational form (once other cues are rendered ambiguous), though residual properties of the base might still be present. We hence propose a differential labeling of the two contours, with continuations being labeled as L*H% and questions as H*H%. Finally, as in the early observations by Grundstrom [8] and Di Cristo (1976), we propose that question rises need to have a large enough F0 span and higher final (perceived) target height, which could be attained by merely modifying the shape of the rise as predicted by the ToG approach.

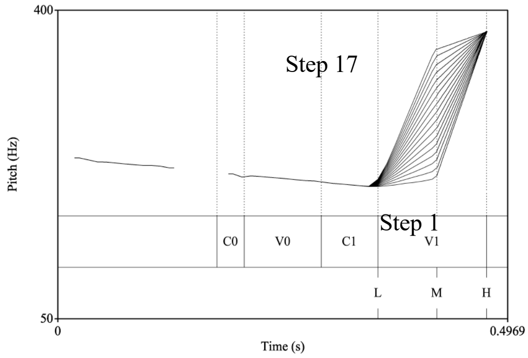


Figure 1: Schema of the 17-step modification with segmental annotations (C for consonant, V for vowel), used to create the 2 continua.

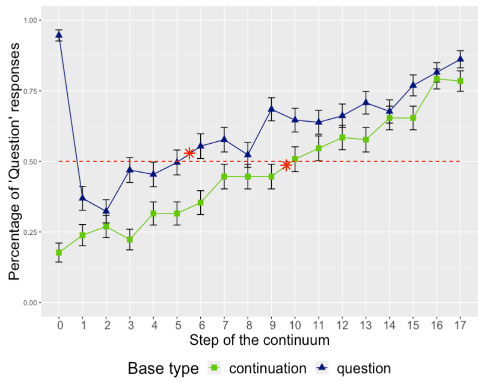


Figure 2: Percentage and standard error of 'question' responses by Step and Base.

REFERENCES

[1] Post, B. (2000). *Tonal and phrasal structures in French intonation*, Ph.D. thesis, The Hague. 583-586.

[2] Delais-Roussarie, E. ; B. Post ; M. Avanzi ; C. Buthke ; A. Di Cristo ; I. Feldhausen ; S-A. Jun ; P. Martin ; T. Meisenburg ; A. Rialland ; R. Sichel-Bazin & H. Yoo (2015). Chapter 3 - *Intonational phonology of French: Developing a ToBI system for French*. In Frota, S & P. Prieto (eds), *Intonation in Romance*. Oxford University Press.

[3] Torreira, F., & Valtersson, E. (2015). "Phonetic and visual cues to questionhood in French conversation". *Phonetica*, 72, 20-42.

[4] Rossi, M., Hirst, D. & Di Cristo, A. (1981). *L'intonation : de l'acoustique à la sémantique*. Paris: Klincksieck.

[5] Valtersson, E., & Torreira, F. (2014). Rising intonation in spontaneous French: How well can continuation statements and polar questions be distinguished? In N. Campbell, D. Gibbon, & D. Hirst (Eds.), Proceedings of Speech Prosody 2014 (pp. 785-789).

[6] Delattre, P. (1966). "Les Dix Intonations de base du français." In *French Review*, 40(1), 1-14.

[7] Di Cristo, A. (1976). Des traits acoustiques aux indices perceptuels : application d'un modèle d'analyse prosodique à l'étude du vocatif en français. TIPA 3 :213-358.

[8] Grundstrom, A. (1973). L'intonation des questions en français standard.Studia Phonetica 8: Interrogation et intonation.19-49

[9] Arvaniti, A., Ladd, D. & Mennen, I. (1998). "Stability of tonal alignment: The case of Greek prenuclear accents." *Journal of Phonetics*, 26: 3-25.

[10] D'Imperio, M. (2012). "Prosodic representations". In A. Cohn, C. Fougeron & M. Huffman (Eds.), *Handbook of Laboratory Phonology*. Oxford: Oxford University Press (section on *Tonal alignment*).

[11] D'Imperio, M. (2000). *The role of perception in defining tonal targets and their alignment,* Ph. D. thesis, the Ohio State University.

[12] Knight, R.A. & Nolan, F. (2006). "The effect of pitch span on intonational plateaux." *Journal of the International Phonetic Association*, 36(1), 21-38.

[13] Niebuhr, O. (2003). *Perzeptorische Untersuchungen zu Zeitvariablen in Grundfrequenzgipfeln.* PhD thesis, U. of Kiel.

[14] Barnes, J., Veilleux, N., Brugos, A. & Shattuck-Hufnagel S. (2012). "Tonal Center of Gravity: A global approach to tonal implementation in a level-based intonational phonology." *Laboratory Phonology* 3(2), 335-383.

[15] Boersma, P. & Weenink, D. www.praat.org (2018). 12983*, 2018. Praat: doing phonetics by computer [Computer program], specified versions 2009.

[16] André C., Ghio A., Cavé C., Teston B. (2003) "PERCEVAL: a Computer-Driven System for Experimentation on Auditory and Visual Perception", Proceedings of XVth ICPhS, Barcelone, Espagne, p. 1421-1424

The effect of instructed second language learning on the acoustic properties of first language speech

Olga Dmitrieva[1], Amy Hutchinson[1], Allard Jongman[2], Joan Sereno[2], Alexis Tews[1]
[1]*Purdue University,* [2]*University of Kansas*

Research demonstrates that second language (L2) learning can affect first language (L1) speech production. This has been shown for proficient long-term second language learners in immigration settings [2] as well as novice learners in the study abroad setting [1]. Moreover, recent studies suggest that immersion situation is not a necessary condition for second language-induced phonetic changes in L1, which can occur under typical instructed L2 learning conditions, whereby learners remain in their first language environment [3, 4]. This evidence suggests that such phonetic changes in L1 may occur in any settings when exposure to and/or use of L2 takes place. It is still not clear however whether some factors may increase or decrease the probability of L2 phonetic influence on L1 speech.

The present paper reports on the results of two production studies exploring the possibility of L1 phonetic drift in the direction of L2 norms in a typical L2 classroom settings in a major Midwestern University. Two group of American students enrolled in intermediate-level Russian (N=18) and French (N=24) language courses were recorded reading a list of Russian and English or French and English words, respectively. The word lists were designed to investigate the acoustic realization of word-initial (French and Russian students) or word-final stop (Russian students only) voicing. Several acoustic correlates of initial and final voicing were measured (VOT, onset f0, preceding vowel duration, closure duration, duration of glottal pulsing during closure, burst duration) and compared to data from two control groups of monolingual speakers of American English (N=18 and N=30) from the same geographic region. Results demonstrated that realization of both initial and final voicing in English of Russian learners were affected by exposure to Russian. In the majority of measures, the drift occurred in the direction of convergence with the norms characteristic of Russian voicing (e.g. Figure 1). Divergence from Russian was detected only in the use of negative VOT for initial voiced stops. No evidence of phonetic drift either towards or away from French norms was detected in the English speech of French learners.

Thus, the results of the studies confirm that L2-triggered phonetic changes in L1 are a possible but not an inevitable outcome of exposure to L2 under the typical conditions of instructed L2 learning. The markedly different results obtained in two experiments conducted in largely equivalent settings, with comparable methodologies and the same population from which the participants were sampled, raise interesting questions about the source of the difference and the underlying causes of the observed L1 phonetic drift. While a definitive answer is not possible without additional research, several possible explanations can be offered.

One possibility is that the differences in the attitudinal, motivational, and cognitive characteristics of the two learner groups can account for the observed divergence in the results. Russian is generally perceived as a more complex target for second language learning than French, thus conceivably attracting more motivated learners who are genuinely interested in the country, its culture, and language.

Another possibility is that the likelihood of L1 phonetic drift is determined by the type of L2 (and possibly L1) input learners receive in the classroom. In the present study, all of the Russian instructors, but not French ones, were native speakers of the target language. Therefore, learners of Russian, but not French, were exposed to native L2 as well as L2-accented English. This suggests that exposure to native L2, rather than the use of L2 per se, is a necessary condition for the phonetic influence of L2 on L1.

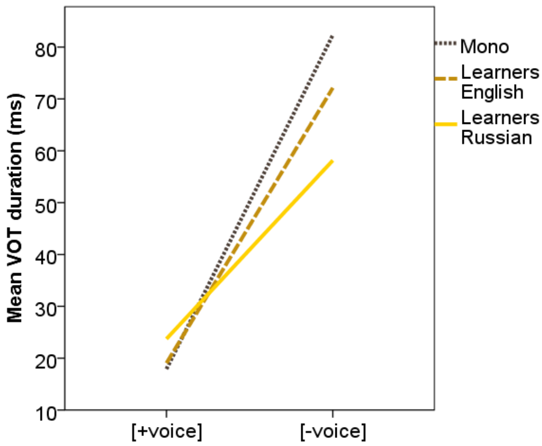


Figure 1. *Voiced and voiceless word-initial stops produced with positive VOT by native monolingual speakers of English (dotted line), learners of Russian speaking English (dashed line), and learners of Russian speaking Russian (solid line).*

[1] Chang, C. B. 2012. Rapid and multifaceted effects of second-language learning on first language speech production. *Journal of Phonetics* 40(2), 249–268.

[2] Flege, J. E. 1987. The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics* 15(1), 47–65.

[3] Herd, W. J., Walden, R. L., Knight, W. L., & Alexander, S. N. 2015. Phonetic drift in a first language dominant environment. *Proceedings of Meetings on Acoustics* 23, 060005.

[4] Huffman, M. K., & Schuhmann, K. S. 2015. Effect of early L2 learning on L1 stop voicing. *Proceedings of Meetings on Acoustics 23*, 060007.

Learning non-native phonological contrasts through multimodal input: the contribution of the tactile sense

Núria Esteve-Gibert[1], Carme Muñoz[2], Natàlia Fullana[2], Ingrid Mora-Plaza[2], Lena Vasylets[2], Joan C. Mora[2]
[1]*Universitat Oberta de Catalunya,* [2]*Universitat de Barcelona*

It is widely accepted in the field of second language acquisition that learners benefit from multimodal information like visual speech gestures (i.e. lip movements) in addition to acoustic information to distinguish and acquire non-native phonological contrasts. A variety of L2 phonetic training studies have shown that listeners improve their perception and production of non-native segmental categories after being trained with audio-visual stimuli (compared to audio-only stimuli) [1-3]. The present study aims at exploring if the multimodal speech information can also be beneficial in the form of tactile cues, and if these tactile cues contribute significantly to the acquisition of L2 phonology.

Previous evidence suggests that tactile information can processed as a source of linguistic information. [4] found that listeners perceive a stop consonant as aspirated if the acoustic input is accompanied by a puff of air applied to the listeners' skin (simulating the burst of the aspiration). Individuals with auditory deficits are also found to benefit from trainings with tactile lip-reading to acquire segmental categories (eg. [5]). Despite these findings, tactile input is far from being a consolidated as a relevant source of phonological information. Our study wants to contribute to this debate by investigating how tactile input can help L2 learners perceiving new phonological contrasts.

We designed two phonetic production training experiments testing two types of tactile input: the production of tactile lip-reading (i.e. placing one's fingers around the lips to 'touch' the lip movements while speaking; in Experiment 1), and the perception of an auditory-tactile association (i.e. touching an object that represents the acoustic properties of a phoneme; in Experiment 2). Experiments are now in the pilot phase and we plan to test 66 L1 Catalan/Spanish speakers in each. Both experiments have a pre-test and a post-test in which discrimination (ABX task) and production (imitation) abilities are evaluated. The stimuli consist of monosyllabic non-words containing a vowel contrast that is non-existent in the participants' L1 (the /æ-ʌ/ contrast). Between the pre- and post-tests participants take part in 3 phonetic training sessions with 3 different conditions (between-subjects): an auditory-only (AO) condition, an audio-visual (AV) condition, and an audio-visual-tactile (AVT) condition. In Exp 1, participants repeat the English native speaker's productions (AO), repeat the target word while looking themselves in a mirror (AV), or repeat the target word while looking themselves in a mirror and touching their own lip movements with their fingers (AVT). No specific instructions of the target lip or tongue positions were given. In Exp 2, what varies is the nature of their exposure to the English speaker: they hear the target words acoustically (AO), they see the native speaker audio-visually (AV), or they see the native speaker audio-visually while holding a specific object with their hands (AVT): a big ball while exposed to /æ/ and a small ball while exposed to /ʌ/. This auditory-tactile association was chosen because a previous test revealed that 38 out of 53 L1 Catalan/Spanish learners of English (a 73%) associate a big ball with the /æ/ phoneme and a small ball is associated with the /ʌ/ phoneme (possibly due to a pairing between object size and vowel length).

The results of these two experiments will reveal if speech information that is perceived through the tactile sense can help L2 learners acquire difficult phonological contrasts, if tactile lip-reading and an acoustic-tactile association are both helpful, and, more generally, if the multimodal nature of speech information can also be extended to the tactile sense.

[1] C. Aliaga-Garcia, "The effect of auditory and articulatory phonetic training on the perception and production of L2 vowels by Catalan-Spanish learners of English," PhD Dissertation. University of Barcelona, 2017.

[2] M. Ortega-Llebaria, A. Faulkner, and V. Hazan, "Auditory-Visual L2 Speech Perception: Effects of Visual Cues and Acoustic-Phonetic Context for Spanish Learners of English," *Speech, Hear. Lang. Work Prog*., vol. 13, p. 3951, 2001.

[3] V. Hazan, A. Sennema, A. Faulkner, M. Ortega-Llebaria, M. Iba, and H. Chung, "The use of visual cues in the perception of non-native consonant contrasts," *J. Acoust. Soc. Am.*, vol. 119, no. 3, pp. 1740–1751, 2006.

[4] B. Gick and D. Derrick, "Aero-tactile integration in speech perception," *Nature*, vol. 462, no. 7272, pp. 502–504, 2009.

[5] G. Plant, J. Gnosspelius, H. Levitt, "The use of tactile supplements in lipreading Swedish and English: a single-subject study," J. Speech. Lang. Hear. Res., vol 43, no. 1, pp. 172-183, 2000.

Acoustic Differences among English -s Allomorphs in a Children's Book Reading Task

Colleen Fitzgerald, Rebecca Ebert, Jason Whitfield
Bowling Green State University, Bowling Green, Ohio, USA

In the acquisition of English morphophonology, the third singular *–s* marker on verbs (e.g., wants) has a protracted course of development compared to other grammatical functions carried by *–s* related morphemes, specifically plural *–s* (e.g., dogs) and possessive *–s* (e.g., Pete's). Phonological constraints cause three allomorphs, or variant surface forms of morphemes, to arise for each of these three morphemes (i.e., [s], [z], and [əz]). Previous explanations for the developmental variation observed for these three morphemes focused on parent input frequency, sentence position, or acoustic duration. In these investigations, the allomorphs of these morphemes were typically collapsed, so it has not been possible to understand interactions between phoneme category (i.e., [s] vs. [z]) and morpheme type (i.e., third singular, plural, and possessive) [1-3]. The current study investigates this interaction in the characteristics of spectral moments for each allomorph production. The two measures reported are the duration of the production and its center of gravity (i.e., mean frequency of spectral energy). If these measures differ between third singular *–s* and the other morpheme types, we could better understand the relative difficulty with third singular *–s* experienced by children acquiring English typically and with language impairment especially.

Participants included ten college students studying speech pathology. Participants were recorded reading a children's book strategically designed to have numerous opportunities for the third singular morpheme, but also included the other *–s* related morphemes. Allomorphs were coded for phoneme category and morpheme type (see Table 1). For the spectral analysis, the audio sample was resampled at 22 kHz, and four equally spaced spectra (i.e., 15 ms window) across the middle 80% of each fricative duration were averaged. The total fricative duration and the center of gravity frequency were computed in Praat [4]. The duration and center of gravity for all 43 allomorph opportunities were averaged across participants. Differences in conditions were analyzed with a linear mixed model.

A significant interaction existed between phoneme category and morpheme type both for duration ($F = 59.35$; $p = .000$) and for center of gravity ($F = 8.85$; $p = .000$; see Tables 2 and 3). First, the phoneme [s] had a longer duration than [z] in monomorphemic words (e.g., "fence" vs. "his") and as a plural marker (i.e., "seats" vs. "dogs"). However, [s] was not longer than [z] when used as a marker of third singular (e.g., "hits" vs. "runs"). Next, the center of gravity was significantly lower in frequency for monomorphemic [s] than for plural [s] ($p = .01$) and possessive [s] ($p = .000$). However, monomorphemic [s] was *not* different from third singular [s].

The majority of English-acquiring children's experience with [s] and [z] occurs in monomorphemic words, in which [s] was significantly longer than [z]. This experience may complicate the acquisition of third singular allomorphs, for which this duration distinction between [s] and [z] was not found. Further complicating the acquisition task is that most of children's experience with [s] as a bound morpheme (i.e., plural *–s* or possessive *–s*) occurs with a significantly different spectral frequency than that of the [s] phoneme in monomorphemic words. Specifically, when marking plural or possessive, the dominant spectral frequency for [s] shifted to a comparatively higher frequency. However, this was not the case when [s] was used to mark third singular, for which center of gravity resembled that of the [s] in single-morpheme words. From infancy, statistical learning supports the identification of morpheme boundaries, so this resemblance could hinder segmentation of the root verb from the third singular marker [5]. Together, these findings contribute to our understanding of why third singular *–s* has a distinctly longer period of acquisition among the s-related morphemes in English.

References

[1] Fitzgerald, C.E., Wettstein, R.C., Ebert, R.L. and Bridges, S.M., 2018. The Phonetic and Linguistic Environment of Third Singular Allomorphs in Parent Input. *Clinical Archives of Communication Disorders*, *3*(2), pp.104-112.

[2] Hsieh, L., Leonard, L.B. and Swanson, L., 1999. Some differences between English plural noun inflections and third singular verb inflections in the input: The contributions of frequency, sentence position, and duration. *Journal of Child Language*, *26*(3), pp. 531-543.

[3] Koehlinger, K., Van Horne, A.O., Oleson, J., McCreery, R. and Moeller, M.P., 2015. The role of sentence position, allomorph, and morpheme type on accurate use of s-related morphemes by children who are hard of hearing. *Journal of Speech, Language, and Hearing Research*, *58*(2), pp.396-409.

[4] Boersma, P., & Weenink, D. 2018. *Praat: doing phonetics by computer*.

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

Table 1

Number of Trials in the Reading Stimulus for each Morpheme (Multiplied by Total Observations across Participants, n=10)

	Phoneme Category	
Morpheme Type	[s]	[z]
Monomorphemic	1 (10)	6 (60)
Plural	2 (20)	1 (8)
Possessive	2 (20)	0 (0)
Third Singular	15 (148)	16 (160)

Note. Values reflect the number of opportunities in the stimulus multiplied by participants (i.e., *n* = 10). One participant did not produce two of the third singular [s] trials. Values with a duration under 35 *ms* were removed for violating the assumptions of the spectral analyses using 15 *ms* windows. As a result, two productions of plural [z] were removed prior to analysis.

Table 2

Mean (Standard Deviation) Duration (ms) of [s] and [z] by Morpheme Type

	Phoneme Category	
Morpheme Type	[s]	[z]
Monomorphemic	154.44(37.66)	78.54(13.21)
Plural	101.37(27.59)	66.06(27.64)
Third Singular	91.40(8.71)	85.40(9.67)

Table 3

Mean (Standard Deviation) Center of Gravity (Hz) for Morphemes by Phoneme Category

		Morpheme Type			
		Monomorphemic	Plural	Possessive	Third Singular
Phoneme	[s]	7442.77(1000.15)	7743.11(1177.99)	8060.94(866.71)	7442.07(945.81)
Category	[z]	7090.37(793.50)	7467.86(1317.20)	--	7091.60(831.79)

Processing prosodic information in sentences with "only" in a second language

Rachida Ganga[1], Haoyan Ge[2], Marijn Struiksma[1], Virginia Yip[3], Aoju Chen[1]
[1]Utrecht University, [2]The Open University of Hong Kong, [3]The Chinese University of Hong Kong

Speakers of West Germanic languages use accentuation to mark new or contrastive information. Usually, misplacement of accent slows down but does not hinder comprehension [1,2]. However, the case is different in sentences with the focus particle "only". "Only" flags upcoming contrasts and accentuation determines the locus of contrast and sentence meaning (e.g. I have only CARRIED the bag vs. I have only carried the BAG). In Dutch, "only" *(alleen)* triggers expectation of adjacent accentuation (early positivity between 100-200ms, followed by an "accent positivity" between 200-500ms) [3]. Nonadjacent accentuation results in reanalysis (P600) [3]. The presence of a linguistic context enhances this processing of accentuation. Corpus studies revealed that "*alleen*" typically precedes the focal word in Dutch even if this separates verbs from their objects [4]. In contrast, in English "only" typically precedes the verb, even when the object is focal [5]. **Do Dutch learners of English use L1 processing strategies when processing focus in L2 English sentences with "only"?** Behavioural research on L2 prosodic processing suggest L1 influence [6] and learner-specific approaches [7], regardless of L2 proficiency. We expect accentuation adjacent to "only" to be similar as L1 processing [3], whereas nonadjacent accentuation will be different. To determine whether Dutch learners of English exhibit L1 processing patterns when processing accentuation adjacent to "only" [3] and learner-specific approach to processing non-adjacent processing in English, we conducted a 64-channel event-related potentials (ERP) study.

Advanced Dutch learners of English (n=33, 14m) listened to English stories (4 types x 60 trials), differing in the presence/absence of context and accentuation on verbs/objects (Table 1). ERPs were analysed from the onset of verbs and objects (t=0) in three time windows (Figure 1) using Mixed Effect Modelling (lme4, R). Baseline correction was done between -100-0ms. ACCENT (verb vs. object), CONTEXT (absence vs. presence), LATERALIZATION (right, middle, left) and ANTERIORITY (front, central, back) are fixed factors and PARTICIPANT and STIMULUS-LIST are random factors. Only effects of ACCENT or interactions with ACCENT will be discussed.

<u>*Adjacent accentuation – verb:*</u> We found the interactions ACCENT x CONTEXT and ACCENT x ANTERIORITY between 100-200ms and ACCENT x ANTERIORITY between 200-390ms (Figure (1A). Accented verbs elicited more positivity, but only with context. This positivity started in frontal and central regions between 100-200ms and spread out to all regions between 200-390ms. In unaccented verbs, context elicited a negativity, suggesting a result of a strong unfulfilled expectancy of accentuation. Thus, context facilitated the expectation and prosodic processing and L2 accentuation elicited cognitive processes that is similar to L1 processing [3].

<u>*Nonadjacent accentuation – object*</u>: We only found an effect of ACCENT (ACCENT x ANTERIORITY) between 500-900ms (Figure 1B). Accented objects elicited a negativity in the frontal and central regions. The lack of ACCENT effects between 100-390ms imply that there is no evidence for processing of the emphatic accent, which is different from L1 Dutch findings for nonadjacent accentuation [3]. Possibly, the expectation for verb accentuation may have been so strong that object accentuation was considered redundant, which goes in line with the "good enough" processing strategy to ease L2 linguistic processing [5]. A sentence reanalysis (P600) only occurred in unaccented objects, different from [3]. Thus, it seems that nonadjacent accentuation in L2 is processed differently than in L1 Dutch [3].

In conclusion, Dutch listeners used similar processing patterns in English as in Dutch [3] when focus was placed on a locus that is preferred in both languages. Focus on a non-preferred locus resulted in processing approach that does not require the processing of accentuation.

Table 1. Examples of experimental stimuli. Pitch accents are represented by capitals.

Context sentences	Target sentences	
The dinosaur has a pumpkin and a bucket. He was going to throw them and kick them. Then he changed his mind.	(A) The dinosaur is only THROWING the bucket.	(B) The dinosaur is only throwing the BUCKET.
-	(C) The dinosaur is only THROWING the bucket.	(D) The dinosaur is only throwing the BUCKET.

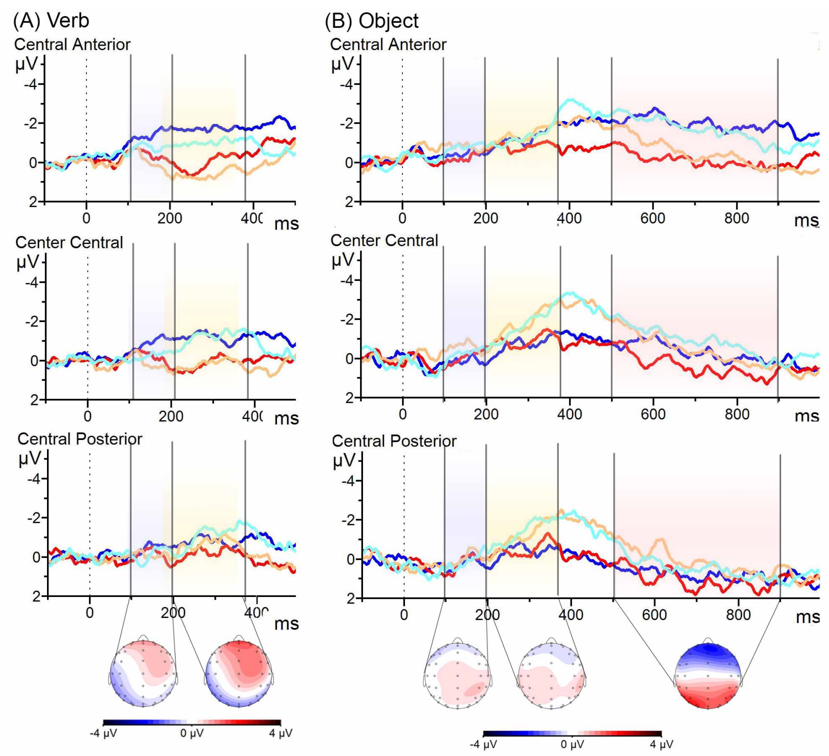


Figure 1. Grand-average ERPs of frontal, central, and posterior of center electrodes for verbs (A) and objects (B) from word onset (t=0). Conditions: accented verbs with context (red), accented verbs without context (orange), accented objects with context (dark blue), and accented objects without context (light blue). Vertical lines indicate boundaries of time windows. We did not analyse the time window 500-900ms at verbs as the ERPs would reflect the processing of words after the offset of verbs. Topographies reflect effect of accent (accented – unaccented condition).

References

[1] Birch, C. & Clifton, C. 1995. Focus, Accent, and Argument Structure: Effects on Language Comprehension. *Language and Speech* 38, 365–391.

[2] Chen, A. 2010. Is there really an asymmetry in the acquisition of the focus-to-accentuation mapping? *Lingua* 120, 1926–1939.

[3] Dimitrova, D. V, Stowe, L. A., Redeker, G., & Hoeks, J. C. J. 2010. Focus Particles and Prosody Processing in Dutch: Evidence from ERPs. In *Speech Prosody* (Chicago, US), 1-4.

[4] Foolen, A., Van Gerrevink, R., Hogeweg, L., & Prawiro-Atmodjo, P. 2009. The placement of focus particles in Dutch. *Linguistics in the Netherlands* 26, 51-63.

[5] Bouma, G., Hendriks, P., & Hoeksema, J. 2007. Focus Particles Inside Prepositional Phrases: A Comparison of Dutch, English, and German. *The Journal of Comparative Germanic Linguistics* 10, 1-24.

[6] Ge, H., Chen, A., Yip, V. 2018. L1 Effects on L2 comprehension of focus-to-prosody mapping: A comparison between Cantonese and Dutch learners of English. In *Speech Prosody* (Poznań, Poland), 187-191.

[7] Chen, A., & Lai, V. T. 2011. Comb or coat: The role of intonation in online reference resolution in a second language. In Zonneveld, W., & Quené, H. (Eds.), *Sound and Sounds.* Utrecht: UiL OTS, 57-68.

Voice patterns associated with age and gender of speakers across the lifespan

Tekla Etelka Gráczi[1,2], Mária Gósy[1,4], Valéria Krepsz[1], Anna Huszár[1], Nóra Damásdi[3], Alexandra Markó[2,4], Ákos Gocsál[1,5]

[1]Research Institute for Linguistics, Hungarian Academy of Sciences; [2]MTA-ELTE Lendület Lingual Articulation Research Group, [3]Eötvös Loránd University, Bárczi Gusztáv Faculty of Special Needs Education, [4]ELTE Eötvös Loránd University, [5]University of Pécs, Faculty of Music and Visual Arts

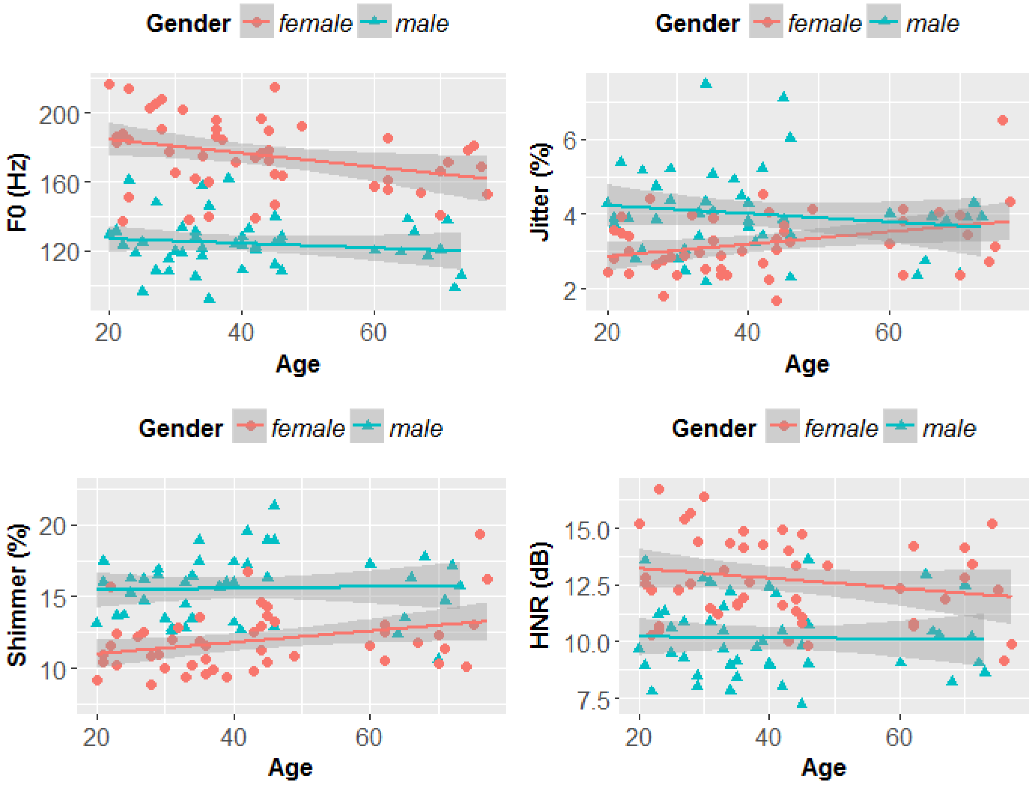
The voice characteristics of speakers are proven to be influenced by various factors, like age and gender due to anatomical, physiological, and social reasons among others [e.g., 1, 2, 3, 4]. Although the presence of these differences are universal due to the anatomical and physiological backgrounds, voice parameters were revealed to have significant differences across languages [e.g., 5, 6, 7]. The present study aims to reveal further data and possible patterns on Hungarian speakers in a cross-sectional study. Our aim is to concentrate on inter- and intraspeaker specific variability.

We hypothesized that (i) and that f0 would show higher values in men's speech and lower values in women's speech of older ages than in the younger speakers' speech. (ii) Mean values and variability of jitter, and shimmer were assumed to be higher with the speakers' age irrespective of their gender, while HNR-values were hypothesized to appear with lower values in elder speakers. (iii) We also assumed that the voice measures in question show gender-specific distributions.

150 healthy speakers (half of them were females) were selected from the BEA database [8]. The ages of the participants ranged from 20 years up to 79 years. 10 men (age: 20, 27, 33, 46, 64, 70 ys) and 4 women (age: 22, 32, 35, 62 ys) were smokers, and 2 men (22, 35 ys) gave up smoking. The database includes various information on the subjects (like age, gender, smoking, weight, etc.), but does not include medicational information (e.g., contraceptives). The material consisted of 25 simple and complex Hungarian sentences to be read aloud by the participants who had ample time to get acquainted with the sentences. All speakers' readings were recorded under the same conditions. The speech samples were segmented into pause-to-pause intervals using the Praat [9]. Measurements of f0, jitter (local), shimmer (local), and HNR were carried out by automatic voice analysis in Praat. The standard settings of f0 range were adjusted according to gender: 50 Hz–350 Hz for males and 75 Hz–450 Hz for females. The values were extracted using a window length of 100 ms with 50 ms overlaps. Linear mixed models were run to analyse the effect of gender, age, and their interaction, while Pearson correlation was used to analyse the age related variance of the measures. Smokers are analysed also whether their results are less specific in their age groups.

The preliminary results (fig. 1) are introduced on a subset of 95 speakers. All parameters confirmed significant differences depending on gender while diverse patterns were found depending on age. The f0 values of female speakers decreased with age while no significant correlation was found with males. The jitter values showed higher values in elder females while no significant difference appeared in male subjects. The values of either shimmer or HNR did not yield any significant correlation with age irrespective of gender.

The expected differences in voice parameters across the lifespan were partly confirmed by our preliminary results. The f0 and jitter values varied along the life span in women, but not in men (see [2]), while the preliminary results for shimmer and HNR were not found to show correlation with age in the present cross-sectional study. The final results on a large database may contribute to mapping the variance of voice parameters.



o

Figure 1. *The average values of f0, jitter, shimmer and HNR across age.*

References

[1] Goy, H., Fernandes, D. N., Pichora-Fuller, M. K., & van Lieshout, P. 2013. Normative Voice Data for Younger and Older Adults. *Journal of Voice* 27(5), 545-555.

[2] Nishio, M., & Niimi, S. 2008. Changes in speaking fundamental frequency characteristics with aging. *Folia Phoniatrica Logopedia* 60(3), 120-127.

[3] Stathopoulos, E. T., Huber, J. E., & Sussman, J. E. 2011. Changes in acoustic characteristics of the voice across the life span: Measures from individuals 4–93 years of age. *Journal of Speech, Language, and Hearing Research* 54, 1011–1021.

[4] Eichhorn, J. T., Kent, R. D., Austin, D., & Vorperian, H. K. 2018. Effects of aging on vocal fundamental frequency and vowel formants in men and women. *Journal of Voice* 32(5), 644.e1-644.e9.

[5] Braun, A.1994. Sprechstimmlage und Muttersprache. *Zeitschrift für Dialektologie und Linguistik* 61,170-178

[6] Wagner, A. & Braun, A. 2003. Is voice quality language-dependent? Acoustic analyses based on speakers of three different language. In *Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS 2003)* 651-654.

[7] van Bezooijen, R. 1995. Sociocultural aspects of pitch differences between Japanese and Dutch women. *Language and Speech* 38(3), 253-265.

[8] Neuberger, T., Gyarmathy, D., Gráczi, T. E., Horváth, V., Gósy, M., Beke, A. 2014. Development of a large spontaneous speech database of agglutinative Hungarian language. In: Sojka, P., Horák, A., Kopeček, I., & Pala, K. (Eds.): *Text, Speech, and Dialogue*. Berlin; Heidelberg; New York: Springer, 424–431.

[9] Boersma, P., & Weenink, D. 2018. *Praat: doing phonetics by computer* [Computer program]. Version 6.0.43, retrieved 8 September 2018 from http://www.praat.org/

Uncovering the geographic origin of immigrant communities:
Transitional gliding in Patagonian Afrikaans

Nicholas Henriksen[1], Lorenzo García-Amaya[1], Andries W. Coetzee[1,2], Jiseung Kim[1], & Daan Wissing[2]
[1]*University of Michigan* & [2]*North-West University (South Africa)*

When examining situations of speech communities originating from immigrants, it is often difficult to determine the exact geographic origins of the initial settlers. This is because use of minority languages in diasporic contexts often does not survive past the third generation. In some cases, historical documents can facilitate the identification of the settlers' regional origins (e.g., Latin-American Spanish; [1]). Our study focuses on the variety of Afrikaans spoken in Patagonia, Argentina. Approximately 600 Afrikaans speakers settled in Patagonia between 1902 and 1906; the historical record regarding the regional South-African origin of these settlers is incomplete and contradictory [2]. However, unlike in many migratory situations, this community remained functionally monolingual in Afrikaans during the first two generations after arrival to Patagonia (until the 1950s). It was not until the mid-twentieth century that the Patagonian region shifted nearly completely to Spanish. The currently oldest speakers are third-generation speakers who acquired Afrikaans as their first language and Spanish as their second language during late adolescence. Due to these rare circumstances, the immigrant language (Afrikaans) survived much longer in this community, thereby affording the opportunity to use linguistic features to make inferences about the South-African regional origins of the first settlers.

Within South Africa, there is documented dialectal and regional variation in Afrikaans. A broad goal of our research is to determine how Patagonian Afrikaans (spoken in Argentina) relates to this dialectal variation. We know that two markers of the major dialectal divide in modern South-African Afrikaans are the presence vs. absence of velar palatalization, as well as [ɛ]-[æ] allophony for /ɛ/ [3]. In this paper we focus on velar palatalization, as a starting point for determining the relative place of Patagonian Afrikaans within the known regional variation in South Africa. Specifically, transitional gliding between /k/ and non-back high and mid vowels occurs in the minority White Afrikaans spoken in the Northern-Cape region of South Africa, but not in the Eastern/Northern regions where so-called "standard" Afrikaans is spoken, yielding respective pronunciation differences like [kjənt] vs. [kənt] for /kənd/ 'child'.

We conducted sociolinguistic interviews with 14 Patagonian Afrikaans (PA) speakers, who live in Argentina, and 11 age-equivalent speakers of "standard" South-African Afrikaans (SAA). From each interview, we labeled all instances of the vocalic portion (i.e., non-back high and mid vowels /iɛə/) following /k/ (n=1498 tokens). In the vocalic portion of every token, we extracted F1 and F2 (Hz) at five equidistant intervals.

The results are summarized in Figures 1 and 2. PA speakers show a slight rise in F1 and a steep fall in F2 across the vowel, indicative of a formant transition from a high front [j] to a lower, more centralized vowel in PA, but not in SAA. We conducted mixed-effects modeling on $\triangle$F1 and $\triangle$F2 (Hz at vowel midpoint - Hz at vowel onset), and found significant effects of speaker group (PA vs. SAA) on both $\triangle$F1 and $\triangle$F2. Figure 3 visualizes this comparison, showing that most PA speakers have larger $\triangle$F1 and $\triangle$F2 than SAA speakers, confirming greater formant movements; this indicates the presence of a transitional glide in PA. Our comparative analysis therefore suggests that the original Patagonian settlers likely came from the Northern-Cape dialect region of South Africa. Altogether, our research shows that linguistic data from isolated immigrant speech communities can sometimes augment incomplete historical records about the exact geographical origin of original settlers.

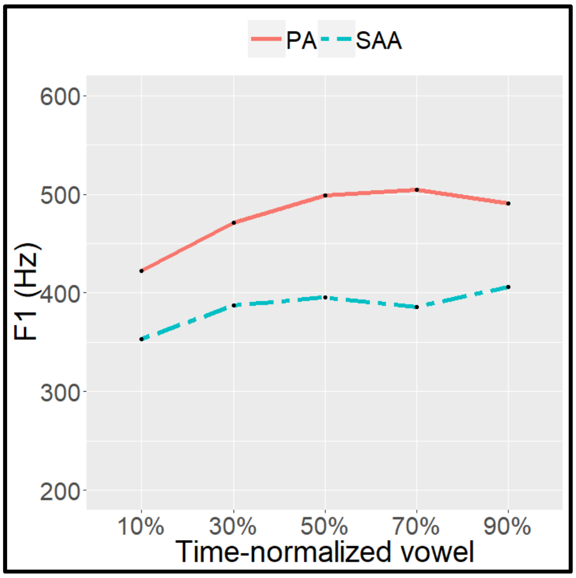


Figure 1 : Average F1

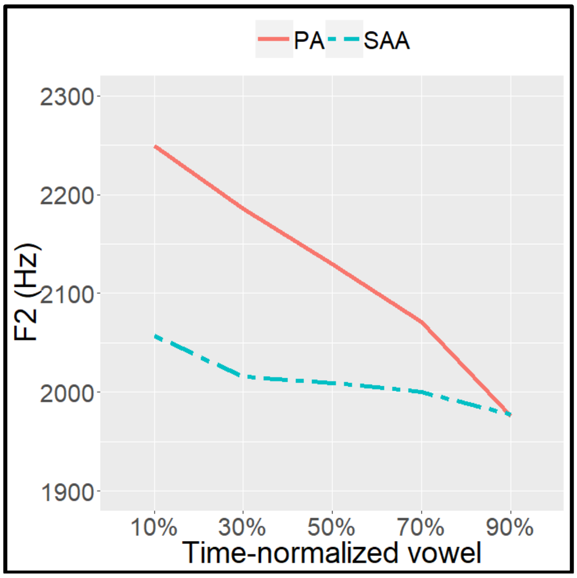


Figure 2 : Average F2

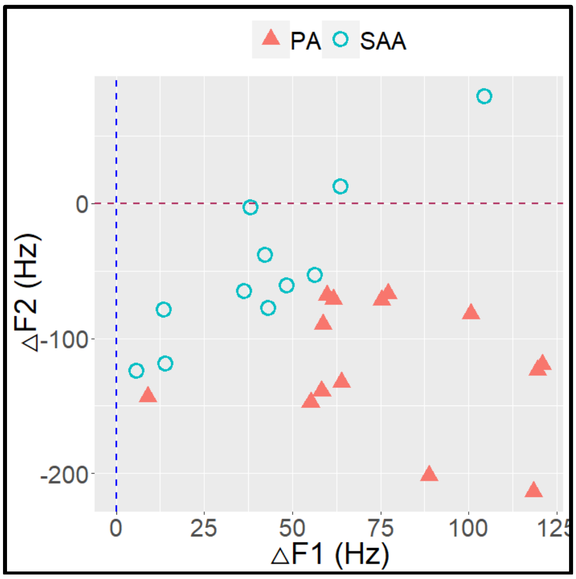


Figure 3 : Individual △F1 and △F2

References

[1] Lipski, J. 1994. *Latin American Spanish.* New York: Longman.
[2] du Toit, B. 1995. *Colonia Boer: An Afrikaner Settlement in Chubut, Argentina*. Lewiston, NY: Edwin Mellen Press.
[3] Roberge, P. 2009. Afrikaans: Considering origins. In Rajend Mesthrie (Ed.), *Language in South Africa* (pp. 79-103). Cambridge University Press.

Intonation affect perlocutionary meaning in requests and offers

Kiwako Ito[1, 2] , James Sneed German[3] , Elisa Sneed German[4] and Caterina Petrone[3]
[1]*University of Newcastle, AUS;* [2]*Ohio State University, US;*
[3]*Aix-Marseille Université, CNRS, LPL, UMR 7309, Aix-en-Provence, France;*
[4]*Université Paul Valéry Montpellier 3, EMMA, Montpellier, France*

Traditional accounts of the semantics of intonational contours assume compositionality, such that the meaning of a given contour depends on the combined functions of pitch accents and boundary tones [1]. This framework, however, has yet to incorporate recent research showing that affective meaning may influence the judgement of speech act (e.g., statement vs. question [2]), that the speaker may choose different tunes (e.g., for requests and offers) according to their familiarity with the listener [3], or that perlocutionary meaning is a function of both sentence type and tune [4].

The main goal of the present research is to explore how perlocutionary meaning is influenced by tune (rising vs. falling) for two distinct, yet comparable illocutionary acts: requests and offers (e.g., *Can [you/I] bring [me/you] some water?*). A main experiment based on a rating task on auditory items was carried out to elicit participants' responses along three scales: speaker MOOD, SINCERITY, and AUTHORITY (cf. [5]). In line with [4], we expected the combination of falling contour and polar question to evoke negative judgments of speaker mood and a perception of higher speaker authority. We were particularly interested in a possible asymmetry between requests and offers with respect to the effects of falling tune on perceived speaker sincerity: speakers can utter offers (*Can I bring you some water?*) without really intending/desiring to carry out the offered act, while by contrast, requests (*Can you bring me some water?*) are unlikely to be produced with no intention/desire of receiving a favor.

Two female native speakers of AE recorded 96 request-offer pairs with both rising and falling contours. Acoustic analyses of the stimuli showed that Speaker 2 had generally larger f0 movements than Speaker 1. In particular, the nuclear pitch accent was higher before falling contours and lower before rising contours, and both the rising and falling contours had larger f0 movements for Speaker 2 than for Speaker 1.

A total of 22677 responses from 237 participants were elicited using a Mechanical Turk online survey. Each participant rated 96 items (6 blocks of 16 items sorted by utterance type (request/offer) and question type (MOOD/AUTHORITY/SINCERITY). Each trial presented an audio file, then a question with a sliding scale (Figure1). Each participant received only one of the three question types (see (1) for an example) per item.

Analyses of the responses confirmed the main effect of falling tune, irrespective of speaker differences, to evoke the perception of a less happy MOOD of the speaker ($t=-16.45$, $p<.001$: Figure2), higher speaker AUTHORITY ($t=8.82$, $p<.001$: Figure3), and less SINCERITY ($t=-11.44$, $p<0.001$: Figure4). In addition, we found that the falling tune raises speaker AUTHORITY to a greater degree for requests than for offers ($t=3.6$, $p<.001$: Figure3) and lowers SINCERITY to a greater degree for offers than for requests ($t=-4.67$, $p<.001$: Figure4).

A follow-up rating task on 52 participants who did not participate in the auditory task was further conducted on the written items. Participants evaluated the difficulty of completing a task or asking someone else to complete a task on their behalf. All items were judged as relatively easy, with requests being judged slightly more difficult than offers ($t=2.8$, $p < .001$). The 'difficulty' score for each item was thus added as covariate in the statistical analyses on the auditory responses of the main experiment. However, no significant effects and interactions were found.

Taken together, these results reinforce findings that intonational tune is a fundamental cue for perlocutionary/affective meaning. In the next steps, we will further test how the presence of the discourse background or knowledge of speaker-listener social relationships influence utterance assessments.

Main experiment

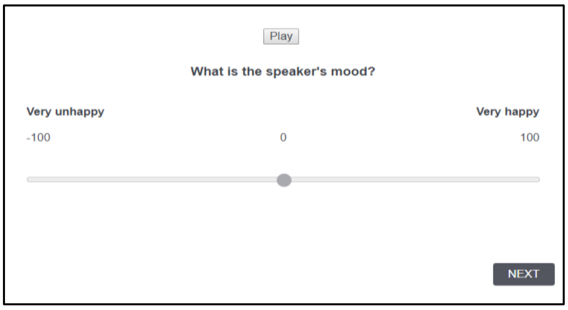


Figure 1. Example Display

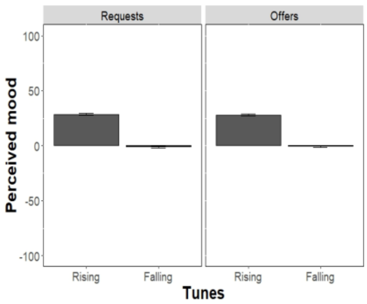


Figure 2. *Mood rating*.

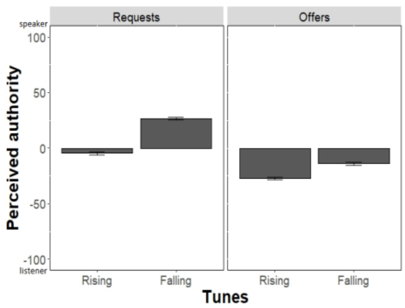


Figure 3. *Authority rating*.

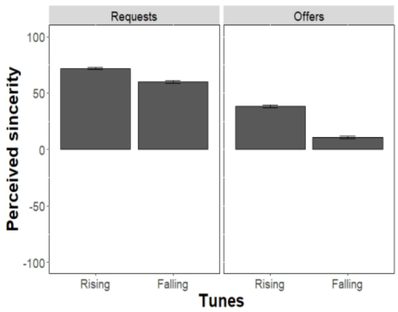


Figure 4. *Sincerity rating*.

(1) *Example target sentence and question set*:

Target request/offer sentence:

Can [you/I] bring [me/you] some water?

MOOD question (for both request and offer):

What is the speaker's mood?

Very unhappy/happy -------------------------------------Very happy/unhappy

AUTHORITY question (for both request and offer):

Who does the speaker think has more authority in this situation?

The speaker/listener ------------------------------------- The listener/speaker

SINCERITY question:

(For request)

Does the speaker want the listener to bring her some water?

(For offer)

Does the speaker want to bring the listener some water?

Not at all/Very much ------------------------------------- Very much/Not at all

References

[1] Pierrehumbert, J., & Hirschberg, J. 1990. The meaning of intonational contours in the interpretation of discourse. In P. Cohen, J. Morgan, & M. Pollack (Eds.), *Intentions in communication* (pp. 342-365). Cambridge: MIT Press.

[2] Pihan, H., Tabert, M., Assuras, S., & Borod, J. 2008. Unattended emotional intonations modulate linguistic prosody processing. *Brain and Language*, 105, 141-147.

[3] Astruc, L., Vanrell, M., and Prieto, P. (2016). Cost of the action and social distance affect the selection of question intonation in Catalan. In: *Intonational Grammar in Ibero-Romance: Approaches across linguistic subfields.* John Benjamins Publishing Company, pp. 91-114.

[4] Jeong, S., Potts, C. (2016). Intonational sentence-type conventions for perlocutionary effects: An experimental investigation. *Proceedings of SALT 26*, 1-22.

[5] Roberts, F., Francis, A.L., Morgan, M. (2006). The interaction of inter-turn silence with prosodic cues in listener perceptions of 'trouble' in conversation. *Speech Communication* 48 (9): 1079–1093.

Final devoicing in the "pool of variation":
A large-scale corpora approach with automatic alignment

A. Jatteau[1], I. Vasilescu[1], L. Lamel[1] & M. Adda-Decker[2]

[1]LIMSI, CNRS, Université Paris-Saclay, Orsay
[2]LPP, CNRS, Université Sorbonne Nouvelle
{jatteau, vasilescu, lamel, madda}@limsi.fr

Research problem. "Final devoicing" (FD) is the process whereby (contrastively) voiced consonants are devoiced in domain-final position (ex. rus. *Youtu*[p]). It is cross-linguistically well attested both as a phonological rule and as a sound change, progressing from larger to smaller domains (phrase to word) [1]. If "sound change is drawn from a pool of synchronic variation" [2], we should then be able to find FD as a variant in languages where it has not been phonologised. The goal of this paper is to investigate this hypothesis in Standard French. French has a voice contrast for obstruents which is maintained in word-final position (e.g. *cage* [kaʒ] "cage" vs. *cache* [kaʃ] "hides"). FD has been reported in French as a regional variant, especially in northern and eastern varieties, in small scale studies of conversational and read speech [3-4]. The present study aims at enlarging the scope of the investigation by quantifying FD in large corpora of Standard French. In this paper we focus on fricatives, which are more likely than stops to undergo devoicing [5], and more specifically on /z/, /ʒ/ and their voiceless counterparts /s/ and /ʃ/. /z/ and /ʒ/ have been reported to be respectively the most and least variable fricative with respect to laryngeal feature variation in word-final position in French [6].

Data. Two manually transcribed corpora were used: ESTER [7], containing 80 hours of semi-prepared speech (radio broadcast news), and the Nijmegen Corpus of Casual French (NCCfr) [8], comprised of 31 hours of face-to-face conversations between friends. These corpora allow us to investigate FD across two different speech styles and several hours of speech.

Methodology. The data was segmented using an automatic speech recognition (ASR) system in forced alignment mode with pronunciation variants [9]. The fricative-final words were extracted, representing 7330 tokens for /z/, 4484 for /ʒ/, 20150 for /s/ and 3000 for /ʃ/ in ESTER, and 1738 for /z/, 1037 for /ʒ/, 4964 for /s/ and 738 for /ʃ/ in NCCFr. The ASR system selected for each fricative among the canonical or (de)voiced variant; for instance /z/ was tagged either [z] or [s]. These data were sorted in 5 classes depending on the following context: whether the next word begins with a voiceless obstruent, voiced obstruent, sonorant, vowel, or if the fricative-final word is followed by a pause (breath or silence).

Results. The results are summarised in Figure 1 and Table 1 below. Two tendencies emerge. The first is laryngeal assimilation: before obstruents, the laryngeal feature of the word-final fricative is sensitive to the laryngeal feature of the following consonant (*mobilise pour* => *mobili*[s]). This tendency is seen for both the voiced fricatives (with a devoicing rate of 66% before voiceless obstruents) and the voiceless ones (voiced in 58% of the cases). These results confirm earlier studies on laryngeal assimilation in French, found in laboratory experiments [6] and in large-scale corpora investigations [7]. The second tendency is FD: /z/ and /ʒ/ are devoiced 24% of the time before a pause, compared to 8% before a vowel and 9% before a sonorant. We can therefore report the presence of FD in Standard French, both in controlled and uncontrolled speech, at a preliminary stage (not in all word-final positions but only before pause). There is no equivalent for /s/ and /ʃ/ (which are voiced 6% of the time before sonorant, 8% before vowel and 6% before pause): as expected from both the typology and the phonetics, there is no "final voicing" in variation in French. The FD effect is reinforced in spontaneous speech: 31% of the final fricatives are devoiced in NCCFr against 22% in ESTER. Finally, we find no effect of place of articulation: in ESTER the devoicing rate for /z/ is 22% vs. 21% for /ʒ/; in NCCFr it is 32% for /z/ vs. 28% for /ʒ/ (not significant). These results are preliminary; in future work we plan to manually check a representative sample of the data to evaluate the accuracy of the automatic segmentation and refine the statistical analyses accordingly.

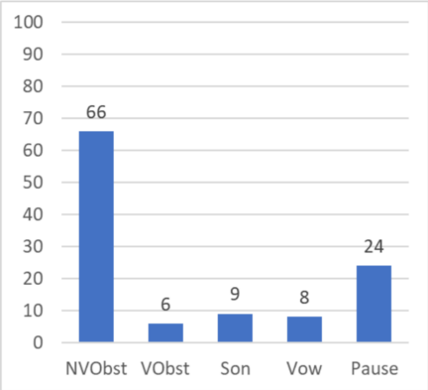


Fig. 1. Devoicing of word-final /z/ and /ʒ/ (%) in ESTER and NCCFr as a function of the following context.
NVObst = voiceless obstruent; VObst = voiced obstruent; Son = sonorant; Vow = vowel.

		Voiceless obstruent	Voiced obstruent	Sonorant	Vowel	Pause	Sum
ESTER	Canonical	876	3 057	1 392	3 170	970	9 465
	Devoiced	1 601	137	102	237	272	2 349
	Sum	2 477	3 194	1494	3 407	1242	11 814
	Canonical %	35	96	93	93	78	80
	Devoiced %	65	4	7	7	22	20
NCCFr	Canonical	275	426	273	662	211	1 847
	Devoiced	649	49	59	78	93	928
	Sum	924	475	332	740	304	2 775
	Canonical %	30	90	82	89	69	67
	Devoiced %	70	10	18	11	31	33

Table 1. Number of occurrences of devoiced /z/ + /ʒ/ across following contexts and corpora.

This work was partially supported by the Maison des Sciences de l'Homme Paris-Saclay as part of a Maturation grant, as well as the French ANR as part of the SALSA project (ANR-14-CE28-0021).

References

[1] Blevins, J. 2006. Theoretical synopsis of Evolutionary Phonology. *Theoretical Linguistics* 32(2). 117–166.

[2] Ohala, J. J. 1989. Sound change is drawn from a pool of synchronic variation. In L. E. Breivik & E. H. Jahr (eds.), *Language Change: Contributions to the study of its causes*, 173–198. Berlin: De Gruyter.

[3] Pooley, T. 1994. Word-final consonant devoicing in a variety of working-class French - a case of language contact? *Journal of French Language Studies* 4(2). 215–233.

[4] Temple, R. 1999. Sociophonetic conditioning of voicing patterns in the stop consonants of French. *Proceedings of the 14th International Congress of Phonetic Sciences.* 1409–1412.

[5] Ohala, J. J. 1997. Aerodynamics of phonology. *Proceedings of the Seoul International Conference on Linguistics*, 92–97. Seoul: Linguistic Society of Korea.

[6] Adda-Decker, M. & P. Hallé. 2007. Bayesian framework for voicing alternation and assimilation studies on large corpora in French. 16th *International Congress of Phonetic Sciences*, 613–616.

[7] Galliano, S., E. Geoffrois, D. Mostefa, K. Choukri, J.-F. Bonastre & G. Gravier. 2005. The ESTER Phase II Evaluation Campaign for the Rich Transcription of French Broadcast News. *Proceedings of ISCA Interspeech*, Lisbon.

[8] Torreira, F., M. Adda-Decker & M. Ernestus. 2010. The Nijmegen corpus of casual French. *Speech Communication* 10(3). 201–212.

[9] Gauvain, Jean-Luc, Lori Lamel & Gilles Adda. 2002. The LIMSI broadcast news transcription system. Speech communication 37(1-2). 89–108.

The Intonational Patterns of Interrogative Sentences in Lithuanian

Asta Kazlauskienė and Sigita Dereškevičiūtė
Vytautas Magnus University, Lithuania

The aim of this study is to identify the main intonational patterns in interrogative sentences in Lithuanian. We constructed 4 types of sentences: a yes/no question without an interrogative word ([ˈgɑːlʲɪmʲɛ ˈɛiˑtʲɪ nɐˈmoː], Eng. *Can we go home?*) and with an interrogative word ([ɐr‿ˈgɑːlʲɪmʲɛ ˈɛiˑtʲɪ nɐˈmoː], Eng. *Can we go home?*), a wh- question ([kɐˈdɐ gɐˈlʲeːsʲɪmɛ ˈɛiˑtʲɪ nɐˈmoː], Eng. *When can we go home?*) and declarative sentence ([ˈgɑːlʲɪmʲɛ ˈɛiˑtʲɪ nɐˈmoː], Eng. *We can go home.*). There are no essential differences between yes/no questions with or without an interrogative word in spoken Lithuanian although a sentence with an interrogative word is more frequent in polite language.

Four male speakers of Standard Lithuanian were asked to read the sentences six times, considering the following context: the participants have been working for a long time and now they are stating that they can go home or asking the head if they can go home or when they will be able to go home. A total of 96 records was obtained. Additionally, a perceptual test was conducted. The interrogative words ([ɐr], [kɐˈdɐ]) were removed, and the participants were asked to identify the type of the sentence: a statement or a question. 21 native speakers (undergraduate students) participated in the experiment. The audio files were annotated using the ToBi labelling system by two phoneticians. It is necessary to note that ToBi for Lithuanian has not been developed yet but there is already research in this area [1]. The empirical data were analysed using PRAAT [2], and graphs were produced using Excel.

The analysis of the intonational patterns allows us to make the conclusion that a statement is characterized by a quite low and down drifting pitch pattern (see Fig. 1 and Table 1). We found two patterns in yes/no questions without an interrogative word. Almost a half of the examples has a rising end, another half has a falling end and an emphasis on the second word. The yes/no questions with an interrogative word have three patterns. One of them has a rising end and an emphasis on the first word, another two patterns are similar in the falling ending but differ in the location of the emphasis. Therefore, we can conclude that an interrogative word is not the most significant indicator of a yes/no question and the emphasized word may allow us to identify these sentences as a question.

Wh- questions can be characterized by two patterns (see Fig. 2) which differ in distribution and prominence of the emphasis. Poor perception of all examples indicates that the interrogative word is the most important indicator of questions in wh- questions.

It should be mentioned that the correlation between the patterns of all interrogative sentences and individual speakers was not observed and examined in the current stage of research.

Summing up, we can draw some conclusions: 1) the end of the interrogative sentence may be rising or falling; if the end is falling, there is an emphasis in the beginning or the middle of a sentence; 2) the theoretically possible variety of intonation contours is reflected in our data base and interrogative sentences can be described by one of the following patterns: a) a rising pitch (%L L* L* H* H%), b) falling-rising (%L H* L* H* H%), c) rising-falling (%L L* H* L* L%), d) falling (%L H* L* L* L% and %L H* H* L* L%). Given the small number of speakers and measurement variables, the results of this study must be taken as provisional.

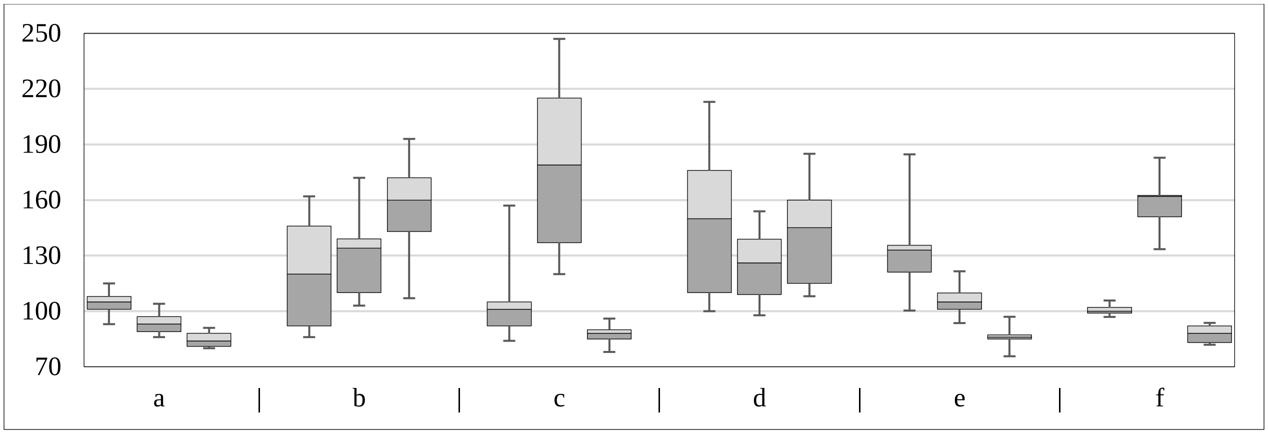


Figure 1. *The scores of F0 mean (Hz): three columns represent three words in sentences; a - a statement, b, c - yes/no questions without an interrogative word (b - %L L* L* H* H%, c - %L L* H* L* L%), d, e, f - yes/no questions with an interrogative word (d - %L H* L* H* H%, e - %L H* L* L* L%, f - %L L* H* L*L%)*

Table 1. *Overview of patterns, frequency and perception of different types of sentences*

Type of sentence	Pattern	Frequency within type of sentence (%)	Accurate perception (%)
statement	%L L* L* L* L% %H H* L* L* L%	98 2	100 100
yes/no question without an interrogative word	%L L* L* H* H% %L L* H* L* L%	56 44	90 94
yes/no questions with an interrogative word	%L H* L* H* H% %L H* L* L* L% %L L* H* L*L%	43 35 22	88 98 91
wh- question	%L H* L* L* L% %L H* H* L* L%	61 39	8 8

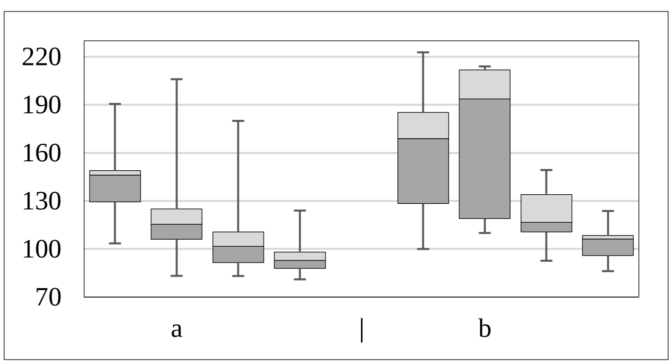


Figure 2. *The scores of F0 mean (Hz) in wh- questions: four columns represent four words in sentences; a - %L H* L* L* L%, b - %L H* H* L* L%*

References:

[1] Hualde, J. I., & Riad, T. 2014. Word accent and intonation in Baltic. *Social and Linguistic Speech Prosody: Proceedings of the 7th International Conference on Speech Prosody* / [ed] N. Campbell, D. Gibbon, D. Hirst, p. 668-67.

[2] Boersma, P., & Weenink, D. 2018. *Praat: doing phonetics by computer*. Computer program.

A corpus-based analysis of Japanese rhythm and mora duration

Sumio Kobayashi [1,2], Amalia Arvaniti[1]
[1]University of Kent [2]Nihon University

A widely held belief, dating back to [1], which was originally published nearly a century ago, is that the duration of moras is held constant in Japanese. This postulated isochrony has given rise to the idea that Japanese is mora-timed [1] and has led to positing a rhythm class of mora-timed languages with Japanese as the prime example. However, acoustic evidence has provided mixed results on this point [2], and has failed to show clear evidence of either isochrony in mora duration or some consistent and generally agreed upon pattern of temporal compensation [see 2 for a review]. A possible reason for these discrepancies could be the reliance of early studies on relatively small samples. The aim of the current study was to examine mora duration using the Corpus of Spontaneous Japanese (CSJ, [3]) in order to uncover possible durational patterns in a large and varied sample of Japanese speech. The CSJ has been compiled by the National Institute for Japanese Language and Linguistics (NINJAL) and includes 50,337 words tagged for accent, mora duration, and mora position in the word.

Mora durations for words of up to 10 moras long were used for analysis. Durations were examined in terms of mora accentuation (accented or unaccented), position in the word (1^{st} to 10^{th}), and segmental composition. In terms of composition a distinction was made between four types, CV moras on the one hand, and single-segment moras, i.e. those consisting of a (coda) nasal (syllable structure CVN), a vowel (syllable structure CVV), and the coda element of a geminate consonant (syllable structure CVG). Durations were analysed using linear mixed models with the lme4 package in R [4, 5]. In all models the dependent variable was duration; mora type, mora position, and accentuation were fixed factors; word was a random factor.

The results showed that mora duration depends on segmental composition, with CV moras being significantly longer than single-segment moras, on average twice as long (see Figure 1a). Accented moras are also significantly longer than unaccented ones (see Figure 1b). Further, mora duration decreases with position in the word, though the effect is small (see Figure 2). Accent and position interact, in that the data show a clear tendency for accent to fall on odd numbered moras: 14.7% of odd-numbered vs. 9.9% of even-numbered moras were accented. This result is related to the fact that even-numbered moras tend to be the second mora of heavy syllables: only 6.6% of odd-numbered moras were single-segment moras, while 30.6% of even-numbered moras were single-segment. A combination of these patterns results in odd-numbered moras being more likely to be longer and accented relative to even-numbered ones.

The results of this large corpus study support the conclusion of [2] and [6] that mora duration is not held constant in Japanese but depends on segmental composition, with CV moras being longer and showing greater variability in duration (see Figure 1a). Further, there is a small but statistically significant effect of accent on mora duration, even though Japanese pitch accent is said not to have durational correlates [6]. Finally, the interaction between accentuation and mora position indicates that the two are connected and that Japanese shows a tendency for odd-numbered moras to be accented. This finding supports accounts like those of [7] and [8] who have argued that Japanese has a tendency for trochaic structure. This tendency is also supported by perceptual evidence, with Japanese listeners showing a preference for the trochaic grouping of auditory stimuli [9].

In conclusion, this extensive corpus does not support the view that Japanese is mora-timed, and indicates that trochees may be the basis of rhythmic structure in Japanese, a tendency that emerges from regularities in the Japanese lexicon and its usage in spontaneous speech. These results add to many studies, such as [10], showing little evidence in favour of isochrony and timing as the basis of rhythm.

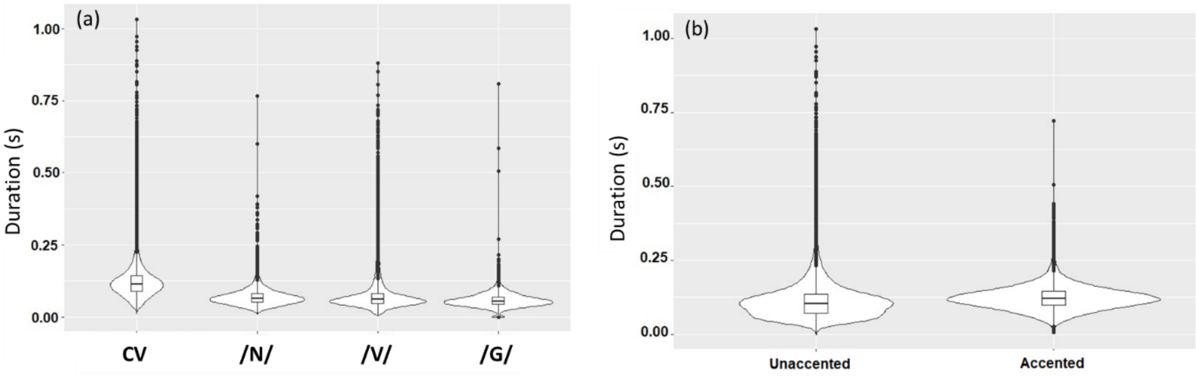


Figure 1. *Duration of moras by segmental composition (a) and accentuation (b).*

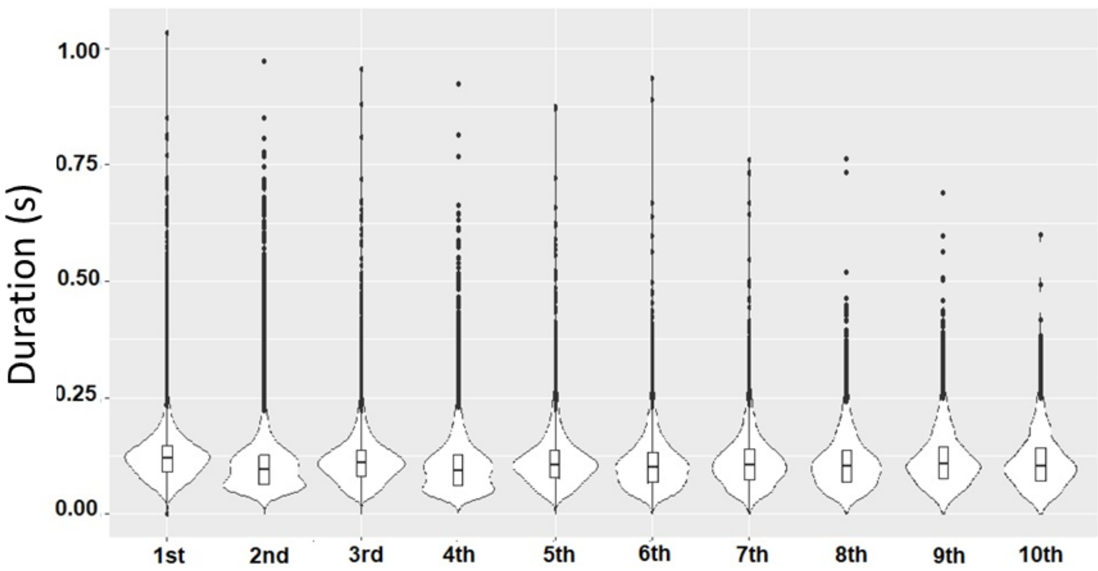


Figure 2. *Duration of moras by position in the word.*

[1] Jinbo, K.1980. Kokugo no onseijou no tokushitsu [The top phonetic characteristics of Japanese]. In T. Shibata, H. Kitamura, H. Kindaichi (eds), *Nihon no gengogaku* (*Linguistics of Japan*), pp. 5-15. Tokyo: Taishukan (originally published 1927).

[2] Warner, N. & Arai, T. Japanese mora-timing: A review. *Phonetica* 58, 1-25.

[3] Maekawa, K., Koiso, H., Furui, S. & Isahara, H. 2000. Spontaneous speech corpus of Japanese. *Proceedings of the Second International Conference of Language Resources and Evaluation* (LREC), volume 2, pp. 947-952.

[4] Douglas B., Maechler, M., Bolker, B., & Walker, S. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1): 1-48.

[5] R Core Team. 2017. *R: A Language and Environment for Statistical Computing*. https://www.R-project.org/.

[6] Beckman, M. E. 1986. *Stress and Non-Stress Accent*. Dordrecht: Foris.

[7] Poser, W. J. 1990. Evidence for foot structure in Japanese. *Language* 66(1): 78-105.

[8] Kitahara, M. 2001. *Category Structure and Function of Pitch Accent in Tokyo Japanese*. Unpublished doctoral dissertation, Indiana University, Bloomington.

[9] Iversen, J. R., Patel, A. D., & Ohgushi, K. 2008. Perception of rhythmic grouping depends on auditory experience. *Journal of the Acoustical Society of America* 124: 2263-2271.

[10] Arvaniti, A. 2009. Rhythm, timing and the timing of rhythm. *Phonetica* 66(1-2): 46-63.

Utterance-final lengthening in infant-directed speech

Anna Kohári[1], Andrea Deme[2,3], Uwe Reichel[1], Ádám Szalontai[1] and Katalin Mády[1]
[1]*Research Institute for Linguistics of the Hungarian Academy of Sciences,* [2]*Eötvös Loránd University,* [3]*MTA-ELTE "Lendület" Lingual Articulation Research Group*

Infant-directed speech (IDS) is among the most commonly investigated types of speech entrainment. It is well known that caregivers speak to their children differently than to adults. Previous research has revealed that higher fundamental frequency, shorter utterances, expanded vowel space, and slower speech rate are typical for infant-directed speech (IDS) compared to adult-directed speech (ADS) [1]. Whether final lengthening is more disproportionally exaggerated in IDS than in ADS relative to the global speech rate still remains a controversial issue. [1-4]. Utterance-final syllables were found to be longer in IDS than in ADS [1, 2]. While some studies have reported significant differences between IDS and ADS in terms of the duration of the phrase-internal syllables or vowels [2, 3], others have not [1, 4]. These results raise the question whether the overall articulation rate differs in the two registers, if the utterance-final syllable or syllables are not taken into account. One study has reported significant differences [2] while another has not [4]. There is rather sparse data available on the temporal features of IDS for Hungarian. A recent study has found no evidence to support the hypothesis that mothers tend to speak slower to their newborn babies [5]. The present study aims to investigate the temporal characteristics of Hungarian infant-directed speech. It is to be tested whether mothers tend to speak slower in general, and whether they lengthen utterance-final syllables even more intensely in IDS than in ADS.

To investigate the features of IDS, recordings of 17 native Hungarian speakers were chosen. Every mother told a story to the experimenter (ADS), then to her 4-month old baby (IDS) using the same story book. The mothers had to improvise a story based on the pictures, but they had to incorporate prescribed utterances word by word. The procedure was repeated when the baby reached the age of 8 months. Utterance boundaries and the last three syllables were annotated in the recordings manually in the Praat 6.0 software [6]. Statistical analysis was carried out in R. We built linear mixed-effect models with directedness (adult vs. infant), and age of infant as fixed effects, and participants and sentences as random effects.

Based on our results, the mothers spoke slower in IDS than in ADS ($p < 0.001$). The last, the penultimate, and the antepenultimate syllable duration of read sentences in IDS differed significantly from that in ADS ($p < 0.01$). Disregarding the last syllable, the speech rate of IDS was markedly slower than that of ADS. The same tendency was found if the last two and even if the last three syllables were excluded from the analysis ($p < 0.01$). Quantifying the lengthening, the duration of the last syllable was found to be around 1.5 times longer than the average duration of all other syllables in the utterance both in IDS and in ADS. The duration ratios of the last two or three syllables compared to the rest of the utterance also appeared to be independent of directedness.

To conclude, IDS is characterized by an overall lower articulation rate compared to ADS, but contrarily to some earlier findings, the final syllables did not extend disproportionally compared to the rest of the utterances in IDS, their boundary marking role was not exaggerated by the mothers. The syllable durations in the utterances were longer in IDS than in ADS, but the timing structure remained the same in terms of duration ratios. Based on earlier experiments [7], it appears that it is the ratio-preserving property of speech that plays a dominant role in the perception of infants, and not the absolute temporal differences. However, the slower speech rate may provide a certain word recognition advantage for the infants when listening to IDS [8]. This study highlights the complexity of the temporal structure in IDS, which can lead to a better understanding of the functions of IDS in language acquisition.

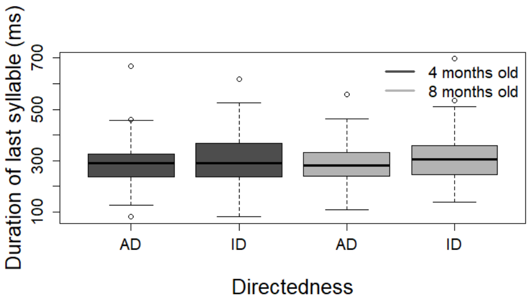


Figure 1. *The last syllable duration in ADS and in IDS at 4 and 8 months of the infant's age*

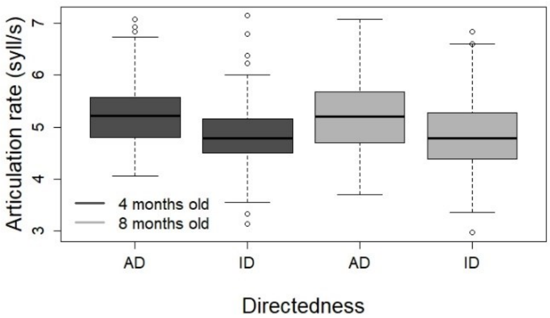


Figure 2. *The articulation rate in ADS and in IDS at 4 and 8 months of the infant's age disregarding the last syllable.*

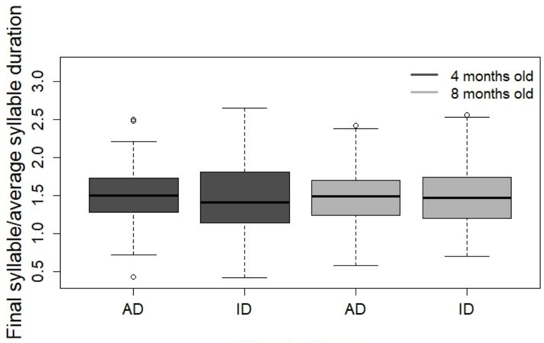


Figure 3. *The last syllable duration divided by the average syllable duration in the rest of the utterance for ADS and IDS at 4 and 8 months of the infant's age.*

This work was supported by the Hungarian National Research, Development and Innovation Office (NKFIH) under grant 'The neurocognitive aspects of early language development' (no. NKFI-115385).

[1] Martin, A., Igarashi, Y., Jincho, N., & Mazuka, R. 2016. Utterances in infant-directed speech are shorter, not slower. *Cognition* 156, 52-59.

[2] Ko, E. S., & Soderstrom, M. 2013. Additive effects of lengthening on the utterance-final word in child-directed speech. *Journal of Speech, Language, and Hearing Research* 56(1), 364-371.

[3] Swanson, L. A., Leonard, L. B., & Gandour, J. 1992. Vowel duration in mothers' speech to young children. *Journal of Speech, Language, and Hearing Research* 35(3), 617-625.

[4] Church, R., Bernhardt, B., Pichora-Fuller, K., & Shi, R. 2005. Infant-directed speech: Final syllable lengthening and rate of speech. *Canadian Acoustics* 33(4), 13-19.

[5] Mády, K., Reichel, U., Szalontai Á., Kohári, A. & Deme, A. 2018. Prosodic characteristics of infant-directed speech as a function of maternal parity In *Proc. 9th International Conference on Speech Prosody 2018* (Poznań, Poland), 294-298.

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

[7] Brannon, E. M., Libertus, M. E., Meck, W. H., & Woldorff, M. G. 2008. Electrophysiological measures of time processing in infant and adult brains: Weber's Law holds. *Journal of Cognitive Neuroscience* 20(2), 193-203.

[8] Zangl, R., Klarman, L., Thal, D., Fernald, A., & Bates, E. 2005. Dynamics of word comprehension in infancy: Developments in timing, accuracy, and resistance to acoustic degradation. *Journal of Cognition and Development* 6(2), 179-208.

Phonetics and distribution of Tourette's verbal tics produced during active speech

Mairym Llorens Monteserín

Persons living with Tourette's syndrome produce an inventory of movements and vocalizations called tics. Tics resemble purposeful, goal-driven movements but are more akin to responses like physiological cough in that they are driven by action-specific urges: producing a tic relieves or reduces the urge to perform that tic (e.g., Jankovic1997, Singer 2005, Felling and Singer 2011). Tic vocalizations that closely mimic linguistic units such as words or phrases are common and are known as verbal tics. Verbal tics are produced on a background of apparently typical voluntary speech. The corpus study presented here includes a battery of linguistically-informed analyses of verbal tics that occurred during active speech, that is, during performance of a talk. First, fundamental frequency is compared across speech and tic modalities as a means to quantify the distinct sources of control over voice these two systems represent. Second, variability in the duration and intonation of the most frequent verbal tic in the corpus is examined, a unique opportunity to observe variability in these parameters in the absence of the intentions to speak that arguably drive their systematic variability involuntary speech. The third analysis is designed to probe one hypothesized linkage between tic and speech motor systems—the extent to which tics are delayed in the service of fluent production of linguistic units like words and phrases. Intentions to speak and urges to tic can be active simultaneously but speech and tics cannot be produced simultaneously. As a result, speech production units on the one hand and tics on the other have to be deployed in some relative order. The distribution of tics relative to the speech around them indexes the extent to which adult ticking "respects" or interferes with speech production. Tic distributional patterning is examined in three ways. First, tics in the corpus are classified according to their temporal distance from adjacent speech and the proportion of tics that are and aren't speech-proximal is computed. Second, the observed likelihood of tics interfering with word production is compared to the likelihood expected by chance if urge-based and goal-driven event timing was occurring independently. Third, the observed likelihood of tics interrupting intonational phrases is compared to the likelihood expected by chance, an analysis mirroring the analysis for word-level plans. Verbal tics are investigated because they are a convenient model urge-based system to compare directly to a goal-driven system—speech.

Voicing, devoicing or contrast enhancement? Russian homorganic nasal-stop sequences in a devoicing context

Mayuki Matsui [1], Silke Hamann [2]
[1]*National Institute for Japanese Language and Linguistics/ JSPS/ UvA,* [2]*University of Amsterdam*

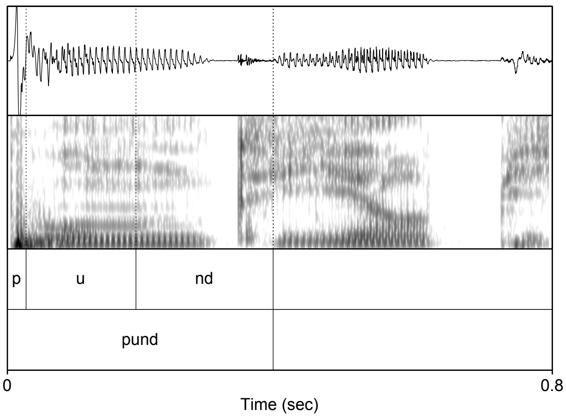
Introduction: Homorganic nasal-stop (NT) sequences have been attracting considerable attention both from phonologists and phoneticians over the decades, since the voicing pattern in NT sequences exhibit cross-linguistic variation. However, little investigated is how the phonetic nature of the NT sequence, which includes a preference for voicing (e.g. Solé 2012) and contrast enhancement (see Hamann & Downing 2017 for an overview), interacts with other phonological/phonetic processes, such as *devoicing* in word-final position. The present study thus examines NT sequences in a *devoicing* context in Russian. By doing this, the present study offers phonetic insights on the current NT typology and on incomplete neutralization in the context of word-final devoicing.

Background: Russian is a language with a two-way voicing contrast in obstruents, and the contrast is maintained in the NT sequences. Previous phonological accounts of Russian have assumed that voiced obstruents are devoiced in word-final position, resulting in neutralization of the voicing contrast (e.g. Halle 1959). However, later careful instrumental studies have shown that even though vocal fold vibration might be absent, other acoustic differences are still distinguishing between "neutralized" obstruents (see e.g. Kharlamov 2014). Similar findings of "incomplete neutralization" have been reported for other languages (e.g. Port & O'Dell 1985 and Roettger et al. 2014 for German). The phonetics of the NT sequence provides insights in the neutralization pattern; if word-final devoicing has priority, the nasal might get devoiced together with the following stop, and vocal fold vibration might cease already during its oral closure. On the other hand, the preceding nasal might block the stop from getting completely neutralized and might itself stay voiced throughout. In addition, the post-nasal voicing contrast might get enhanced by utilizing other phonetic cues, including a long-distance coarticulatory effect of voicing on the vowel duration (see e.g. Chen 1970). An acoustic study was conducted to test these possibilities.

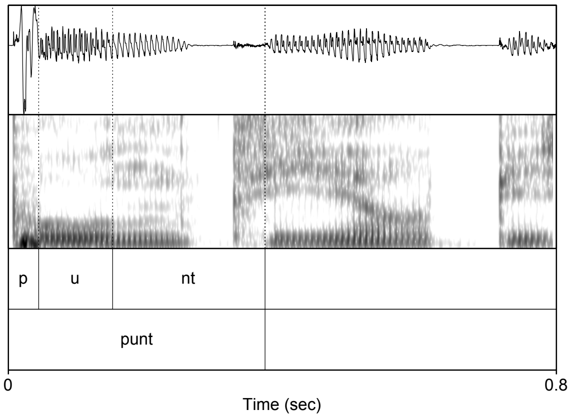
Data: 11 native Russian speakers produced nonce words containing /nd/ and /nt/ sequences in an utterance-medial word-final position (e.g. /pu**<u>nd</u>**/ vs. /pu**<u>nt</u>**/), followed by a sonorant ([j]). The speech materials were elicited using a declension task with auditory stimuli (see e.g. Roettger et al. 2014 for a similar methodology). We measured temporal properties, *inter alia,* the duration of the preceding vowel, nasal and non-nasal, oral closure, and burst. Additionally, the harmonics-to-noise ratio (HNR) of each interval was considered to assess contrast enhancement along the spectral dimension (see e.g. Downing & Hamann 2018 for Tumbuka).

Results and Discussion: Our findings include that (i) the post-nasal stops showed partial voicing for /nd/ but also for /nt/ (Figs. (a) and (b) respectively); (ii) the sonorants (vowel + nasal) preceding /d/ were longer than those preceding /t/ (compare Figs. (a) with (b)), showing incomplete neutralization. However, the vowel itself did not show significant difference, suggesting that a long-distance coarticulatory effect of the final stop on the vowel is questionable in a devoicing context. (iii) The HNR trajectories of the nasal interval were different, depending on the voicing in the following stop. Taken together, the results offer a complex case of incomplete neutralization, interacting with competing phonetic constraints for the NT sequence. The broader implications of these findings for speech perception and NT typology will be addressed during the presentation.

Figures



(a) An example of the word-final (non-utterance-final) /nd/ *sequence in* /pu**nd**/ [punt].



(b) An example of the word-final (non-utterance-final) /nt/ *sequence in* /pu**nt**/ [punt].

References

Chen, M. 1970. Vowel length variation as a function of the voicing of the consonant environment. *Phonetica* 22: 129–159.

Downing, L. J., Hamann, S. 2018. The phonetics of NCh in Tumbuka and its implications for diachronic change. *Papers in Historical Phonology* 3: 77–95.

Halle, M. 1959. *The Sound Pattern of Russian*. Hague: Mouton.

Hamann, S., Downing, L. 2017. *NT revisited again: An approach to postnasal laryngeal alternations with perceptual cue constraints. *Journal of Linguistics* 53: 85–112.

Kharlamov, V. 2014. Incomplete neutralization of the voicing contrast in word-final obstruents in Russian: Phonological, lexical, and methodological influences. *Journal of Phonetics* 43: 47–56.

Port, R., O'Dell, M.L. 1985. Neutralization of syllable-final devoicing in German. *Journal of Phonetics* 13: 455–471.

Roettger, T. B., Winter, B., Grawunder, S. 2014. Assessing incomplete neutralization of final devoicing in German. *Journal of Phonetics* 43: 11–25.

Solé, M-J. 2012. Natural and unnatural patterns of sound change? In M-J. Solé and D. Recasens (eds.), *The Initiation of Sound Change: Perception, Production, and Social Factors*, 123–145. Amsterdam: John Benjamins.

To what extent the French prosodic encoding of contrast is addressee-oriented?

Amandine Michelas[1] and Maud Champagne-Lavau[1]
[1] *Aix Marseille Univ, CNRS, LPL, Aix-en-Provence, France*

According to [1] speakers use intonation to specify a relationship between the content of a phrase and the mutual knowledge/belief of participants in the current discourse about this content. For instance, in American English, the L+H* pitch accent is used by speakers to mark a correction or a contrast. In this case, by using this pitch accent, the speaker signals that the accented item and not some alternative related item should be mutually believed, all the alternatives being shared by both the speaker and the addressee. This framework thus assumes that the speaker's prosodic encoding of contrast depends i) on the referential context itself and ii) on what the speaker and the addressee know about this context. Following this idea, lots of research has focused on the description of the prosodic forms involved in contrast encoding in different languages. However, most studies have not investigated to what extent the prosody of contrast reflects the way speakers take their addressees into account. Do prosodic choices of speakers mainly reflect the referential context from the speaker's point of view independently of his/her addressee's presence and knowledge or does the speaker prosodically encode contrast on the basis of shared knowledge to serve the addressee's needs?

To disentangle between these two possibilities, we investigated whether the French prosodic encoding of contrast is affected by the presence of an addressee. The main difference between French and American English regarding contrast encoding is that French speakers don't use pitch accent type to signal the contrastive status of a referent. Rather, they can use a large variety of strategies among which prosodic phrasing appears as one of the most commonly used strategies [2,3,4,5]. For instance, in noun-adjective pairs such as *bougies violettes* vs. *bonbons violets* 'purple candles' vs. 'purple candies', French speakers parse the noun in the 2nd fragment in a separate prosodic phrase from the following adjective when this noun contrasts with the 1st noun in the pair (e.g., *bougies violettes* followed by [*BONBONS] [violets]*). By contrast, they produce it in the same prosodic phrase when it refers to the same object but with a different modifier *(bonbons marron* 'brown candies' followed by [*bonbons violets*] 'purple candies').

In this study, 30 native speakers of French played an interactive game developed by [5]. During this game, participants had to indicate a given route from a departure point to an arrival point by producing noun-adjective pairs in which the noun in the 2nd noun-adjective fragment (the target noun) was either identical to the noun in the 1st fragment (e.g., *bonbons marron* 'brown candies' vs. *bonbons violets* 'purple candies') or contrasted with it (e.g., *bougies violettes* 'purple candles' vs. *BONBONS violets* 'purple candies'). We also manipulated the presence vs. absence of an addressee meaning that 15 participants performed the task with an addressee whereas the other 15 described the route while no addressee was present, and no potential addressee was mentioned in the instructions. Prosodic phrasing produced by participants was measured in terms of whether the target noun was phrased within the same Accentual Phrase as the following adjective (1-AP phrasing) or whether it was phrased in a separate AP (2-AP phrasing). Results confirmed those of Michelas et al. (2014) showing that speakers produced more 2-AP phrasing when the target noun was contrastive in the presence of an addressee (Figure 1). By contrast, in the absence of an addressee, speakers did not produce more 2-AP phrasing than 1-AP phrasing meaning they did not use prosodic phrasing to encode the contrastive status of target nouns. Our results are difficult to reconcile with the view that prosodic encoding of focus reflects the referential context independently of the presence of an addressee. Rather it appears that in a language such as French in which prosodic phrasing is the most common strategy to encode contrast, parsing choices reflect the way speakers take their interactional partner into account.

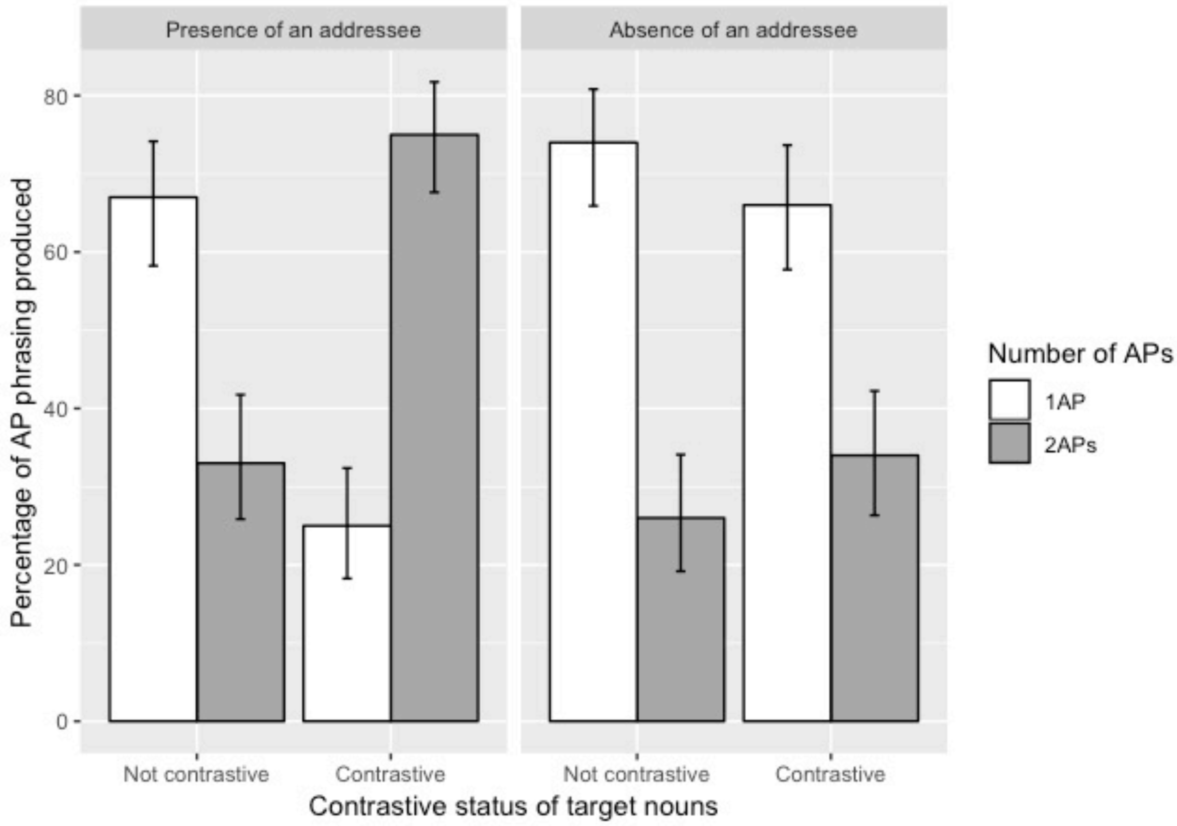


Figure 1. *Percentage of prosodic phrasing produced by participants depending on the number of APs they produced (1 AP vs. 2 APs) and the contrastive status of target nouns (not contrastive vs. contrastive). Error bars show a default 95% confidence interval.*

[1] Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. Cohen, J. Morgan, & M. Pollack (Eds.), *Intentions in communication* (pp. 342–365). Cambridge: MIT Press.

[2] Féry, C. (2001). "Intonation of focus in French". in *Audiatur Vox Sapientes: A Festschrift for Arnim von Stechow,* ed. C. Féry, W. Sternefeld (Berlin: Akademi Verlag), 153-181.

[3] Dohen, M., & Lœvenbruck, H. (2004). Pre-focal rephrasing, focal enhancement and postfocal deaccentuation in French. in *Proceedings of Interspeech 2004.*

[4] Chen, A., and Destruel, E. (2010). Intonational encoding of focus in Toulousian French, in *Proceedings of Speech Prosody 2010.*

[5] Michelas, A., Faget, C., Portes, C., Lienhart, A.-C., Boyer, L., Lançon, C., & Champagne-Lavau, M. (2014). Do patients with schizophrenia use prosody to encode contrastive discourse status?. *Frontiers in Psychology*, 5:755.

Effects of L2 Experience on the Realization of L1 Phonological Neutralization: Incomplete Devoicing in Bulgarian-English Bilinguals

Sejin Oh[1,2] Jason Bishop[1,3] Chen Zhou[1]

[1]*CUNY Graduate Center* [2]*Haskins Laboratories, CUNY* [3]*College of Staten Island, CUNY*

A large body of work has investigated the extent to which word-final devoicing, in languages that feature this phonological process, is phonetically complete ([1–9]). Where incompleteness occurs, it takes the form of small durational differences in the direction of the underlying contrast—for example, shorter closure duration for the final obstruent and longer duration for the vowel preceding it, in German /ʁaːd/ *'wheel'* versus /ʁaːt/ *'advice'* ([6,10]). Though small, the differences in question have been found statistically significant in the majority of studies with enough power to reasonably detect them ([10,11]). In the present study we investigate the phonetic realization of final devoicing in Bulgarian, which, to our knowledge, has not been reported on in the incomplete neutralization literature previously.

In addition to confirming the existence of incomplete neutralization in this language, we sought to explore the extent to which it is sensitive to speakers' contact with English, a language lacking final devoicing. Although researchers have often noted the likely relevance of second language (L2) experience in incomplete neutralization ([5,12]) we are aware of only one study that investigated it systematically. In particular, [13] compared the productions of Russian monolinguals with Russian(L1)-English(L2) bilinguals and found that, while multiple acoustic measures showed monolingual Russian speakers' final devoicing to be incomplete, native Russian speakers with proficiency in L2 English showed even less completeness. In our study, we compared the productions of native Bulgarian speakers residing in Bulgaria versus those residing in the US.

Data come from a production study that is part of a larger investigation into phonological neutralization in Bulgarian. Thirty-four native Bulgarian speakers (half residing and recorded in the United States, half in Bulgaria) produced 13 minimal pairs contrasting in the voicing status of a final stop or fricative (e.g., /rɔg/ *'horn'*, /rɔk/ *'rock'*; /kub/ *'cube'*, /kup/ *'bunch'*; /ʃɛv/ *'seam'*, /ʃɛv/ *'boss'*). Speakers read 3 repetitions of a stimulus list presented sequentially on a computer screen, on each trial reading the item both in isolation and in a carrier sentence. The following acoustic measurements were taken for targets words: duration of the final obstruent's closure/frication; duration of voicing into closure/frication (calculated as a proportion of the obstruent's duration); duration of the vowel preceding the final obstruent; and the duration of a release burst (for final stops, when present). As analysis is still in progress, we preview patterns rather than statistical results, and do so for a subset of the data (N=6 out of the 17 speakers recorded in Bulgaria; 17 out of 17 of the speakers recorded in the US).

First, although still exploratory at this point, durational patterns consistent with incomplete neutralization were apparent for speaker groups in both locations, and of the small magnitude that is typically reported in studies of this phenomenon. Speakers residing in the US, however, showed less complete neutralization than speakers residing in Bulgaria, on three of the four acoustic measures—all but release burst duration. Second, among speakers residing in the US, the length of residence was correlated with two measures of incompleteness, namely preceding vowel duration and the duration of the final obstruent itself (Figure 1). Notably, these results very much resemble the patterns reported for Russian incomplete neutralization in [13]. We will discuss the implications of such findings for theories of the mechanism underlying subphonemic differences like incomplete neutralization, which have been argued to reflect lexical activation rather than a functional/communicative function (e.g., [14]).

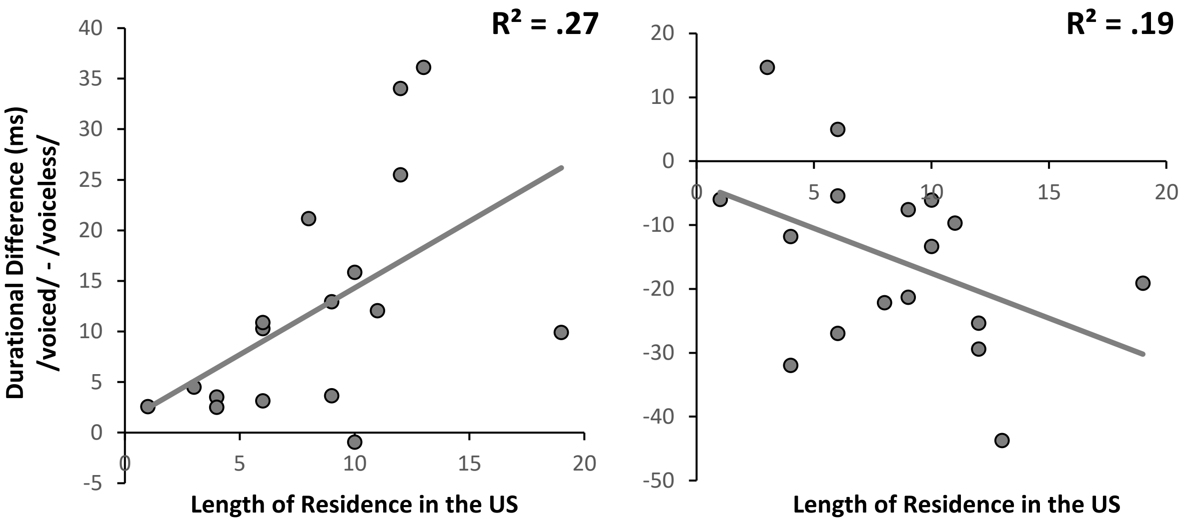


Figure 1. *Degree of incomplete neutralization as a function of residency in the US: Mean differences in preceding vowel duration (left) and final obstruent duration (right) for tokens with underlyingly voiced final obstruents relative to those with final voiceless ones.*

[1] Mitleb, F. (1981). Temporal correlates of "voicing" and its neutralization in German. *Research in Phonetics* 2, 173–191. (Bloomington, Indiana: Indiana University).

[2] Fourakis, M., & Iverson, G.K. (1984). On the 'incomplete neutralization' of German final obstruents. *Phonetica* 41, 140–149.

[3] Dinnsen, D.A., & Charles-Luce, J. (1984). Phonological neutralization, phonetic implementation and individual differences. *Journal of Phonetics* 12, 49–60.

[4] Port, R., & O'Dell, M. (1985). Neutralization of syllable-final voicing in German. *Journal of Phonetics* 13, 455–471.

[5] Port, R., & Crawford, P. (1989). Incomplete neutralization and pragmatics in German. *Journal of Phonetics* 17, 257–282.

[6] Piroth, H.G., & Janker, P.M. (2004). Speaker-dependent differences in voicing and devoicing of German obstruents. *Journal of Phonetics* 32, 81–109.

[7] Warner, N., Jongman, A., Sereno, J., & Kemps, R. (2004). Incomplete neutralization and other sub–phonemic durational differences in production and perception: Evidence from Dutch. *Journal of Phonetics* 32, 251–276.

[8] Kleber, F., John, T., & Harrington, J. (2010). The implications for speech perception of incomplete neutralization of final devoicing in German. *Journal of Phonetics* 38, 185–196.

[9] Winter, B., & Roettger, T.B. (2011). The nature of incomplete neutralization in German. *Grazer Linguistische Studien* 76, 55–74.

[10] Roettger, T.B., Winter, B., Grawunder, S., Kirby, J., & Grice, M. (2014). Assessing incomplete neutralization of final devoicing in German. *Journal of Phonetics* 43, 11–25.

[11] Kirby, J., & Sonderegger, M. (2018). Mixed-effects design analysis for experimental phonetics. *Journal of Phonetics*, 70, 70–85.

[12] Jassem, W., & Richter, L. (1989). Neutralization of voicing in Polish obstruents. *Journal of Phonetics* 17, 317–325.

[13] Dmitrieva, O., Jongman, A., & Sereno, J. (2010). Phonological neutralization by native and non–native speakers: The case of Russian final devoicing. *Journal of Phonetics* 38, 483–492.

[14] Ernestus, M., & Baayen, R.H. (2006). The functionality of incomplete neutralization in Dutch: The case of past-tense formation. In: L.M. Goldstein, D. H. Whalen, & C.T. Best (Eds.), *Laboratory Phonology* 8. (pp. 27–49). Berlin: de Gruyter.

Conflicting forces of place and manner of articulation:
A reaction times study on Polish phonotactics

Paula Orzechowska
[1]*Adam Mickiewicz University, Poznań*

This paper discusses online processing of word-initial phonotactics in Polish. Cross-linguistic research on consonant clusters has demonstrated that online processing is largely sensitive to sonority violations (e.g. Pitt 1998, Moreton 2002, Berent et al. 2007) and place constraints (e.g. Frisch & Zawaydeh 2001). With these views in mind, words differing in terms of the place of articulation and manner of articulation properties are expected to be processed differently. In Polish, the processing of sonority was investigated in terms of ERPs (Wagner et al. 2012, Wiese et al. 2017). However, no studies we are aware of have tested the online processing of consonant clusters in Polish explicitly in terms of reaction times (RTs), and – what os more – using constraints related to the place of articulation.

The experiment explores the psycholinguistic reality of three factors which can be potentially relevant in the processing of word-initial consonant clusters: (1) *cluster existence*, (2) *phonological well-formedness* and (3) *phonetic distance*. Existence distinguishes between clusters that are found in real (existent) or hypothetical words (non-existent). Well-formendness specifies whether CCs follow the *Sonority Sequencing Generalization* (well-formed) or violate it (ill-formed). Examples of clusters used as stimuli are presented below.

	Existent	Non-existent
Well-formed	tr lj gm	nr ʤm ʦx
Ill-formed	ɕʨ lv sk	mz lk jm

Additionally, for each of the conditions, target CCs were further subcategorized depending on the distances in place of articulation between constituent consonant along the following 7-point scale: bilabial–labio-dental–dental–alveolar–alveolo-palatal–palatal–velar. The distance of 0 holds between consonants sharing the place features (e.g. /dl ɕʨ/), while the distance of 6 specifies extreme articulations (e.g. /gm kp/). Target clusters were embedded in monosyllabic nonce words of the structure CCVC, where VC suffixes involved /es/, /ot/, /um/. The total of 252 nonce stimuli were presented to subjects auditorily. 38 native speakers of Polish (average age: 21, 33 women) took part in the experiment. The subjects were requested to indicate whether the words they heard sounded as if they could exist in Polish by pressing a 'yes' or 'no' keyboard button. The data was recorded with the *E-Prime* software.

The statistical analysis involved running the quantile regression (Fasiolo et al. 2017). Two types of data were analysed. First, response latencies are affected by the place of articulation distances. The processing of clusters with medial distances (2–4) is cognitively most costly (the longest reaction times). Larger distances (5-6) entail shorter RTs, and therefore facilitate processing. Interestingly, no statistically significant results were obtained existence and sonority. The study demonstrates that speakers are sensitive neither to the distinction between words with real and hypothetical consonant sequences nor to sonority violations. However, the findings lend support to the principle of the clarity of perception according to which contrast facilitates perception. As regards behavioural data, accuracy rates are affected by the interaction between sonority and existence. This result suggests that sonority is consulted when intuitive judgements is made on new words.

Overall, the results show that although sonority is evoked by native speakers of Polish in the decision making process, the processing of phonotactics is governed by principles other than sonority, i.e. related to the place of articulation. Generally, we conclude that sonority (based on the manner features) does not suffice as a generalizing principle in a phonotactically elaborate language such as Polish.

[1] Berent, I., D. Steriade, T. Lennertz, and V.Vaknin. 2007. What we know about what we have never heard: Evidence for perceptual illusions. *Cognition* 104 (3): 591-630.

[2] Fasiolo, M., Y. Goude, R. Nedellec, and S. N. Wood. 2017. Fast calibrated additive quantile regression . Manuscript. Bristol: University of Bristol.

[3] Moreton, E. 2002. Structural constraints in the perception of English stop-sonorant clusters. *Cognition* 84 (1): 55-71.

[4] Pitt, M. 1998. Phonological processes and the perception of phonotactically illegal consonant clusters. *Perception and Psychophysics* 60 (6): 941-951.

[5] Wagner, M., M. Brett, V. L. Shafer, and M. Steinschneider. 2012. The phonotactic influence on the perception of a consonant cluster /pt/ by native English and native Polish listeners: A behavioral and event related potential (ERP) study. *Brain and Language* 123: 30-41.

[6] Wiese, R., P. Orzechowska, P. Alday, and C. Ulbrich. 2017. Structural principles or frequency of use? An ERP experiment on the learnability of Polish consonant clusters. *Frontiers in Psychology – Auditory Cognitive Neuroscience* 7: 2005.

Assessing pragmatic prosody in 3- to 4-year-old children

Mariia Pronina[1], Iris Hübscher[1,2], Ingrid Vilà-Giménez[1] and Pilar Prieto[3,1]
[1]*Universitat Pompeu Fabra,* [2]*University of Zurich,* [3]*Institució Catalana de Recerca i Estudis Avançats*

Prosodic and pragmatic abilities increasingly develop during preschool years and beyond. However, while some prosodic tests have been developed for children (e.g., PROP [1], PVSP [2], DANVA 2 [3], PEPS-C [4], PPAT [5], MNTAP [6]), they primarily serve for clinical use and for an assessment of children with atypical language development. Moreover, they focus principally either on receptive prosodic skills, or on very basic expressive prosodic skills and do not fully integrate the pragmatic functions of prosody. While recent studies demonstrate that prosodic and pragmatic abilities develop hand in hand (see [7], for a review), to our knowledge, none of the available assessment tools consider these abilities in an integrated way in children.

In this study we present a new Audiovisual Pragmatic Test (APT) to jointly assess prosodic and pragmatic abilities in typically developing Catalan-speaking children. The elicitation methodology is based on the Discourse Completion Task. Whereas the DCT is a widely used method for assessing adult intonational grammar and it has been shown to be an applicable instrument to obtain semi-spontaneous data in adults (see [8], for a review), to our knowledge it has not been widely used to assess prosodic development in children. The APT consists of a total of 47 items social scenarios accompanied by images, which require children to produce speech appropriate to the given situation (see Fig 1). For the purposes of the study, given that the test- takers would be 3- to 4-year-old children, only the first 35 items of the test were applied. While some selected items have been extracted from adult DCT questionnaires on Catalan prosody, the main pragmatic coverage has been based on previous pragmatic tests designed for children (e.g., CASL-2 [9], CELF-5 [10], TOPL-2 [11]). The present study aims to investigate (a) whether this elicitation methodology is adequate for use with young children and (b) which areas of pragmatic prosody have been acquired by the age of 3-4 years. One hundred 3- to 4-year-old Catalan-speaking children participated in the study. For each item, a prosodic score and a pragmatic score were given. The prosodic score reflected the child's ability to produce contextually adequate pragmatic prosody and to use direct speech in first-person to answer each item (i.e., to enact the response); the pragmatic score depended on the adequacy, complexity and specification of the response. First, results indicated that the test was adequate and feasible for 3- to 4-year old children. From the total amount of 3500 potential responses (35 items × 100 children), 49,9% (1748 target sentences) of pragmatically appropriate responses and 36,9% of prosodically appropriate responses (1294 target sentences) were obtained. A large majority of the children engaged in the activity to one degree or another, with only 1% of the group failing to enact any item. These results indicate that the APT instrument allows to gather a considerable amount of intonation patterns produced by preschool children starting at three, which confirms the sensitivity and suitability of the measure for the youngest children. Second, results indicate that 3- to 4-year-old children successfully produced non-biased statements and questions, as well as imperatives and vocatives, and had more difficulties with biased sentences (e.g., with expressing uncertainty, incredulity, surprise, obviousness, corrective and contrastive focus, among others).

Overall, we regard this test as a valuable tool for the assessment of prosodic and pragmatic features in language development. Finally, by presenting this tool, we aim at promoting the inclusion of pragmatics as a relevant dimension in the prosodic assessment practice.

(1) Example item of the APT showing text and illustration.

Figure 1. *Expressing of concern for a friend: "Acabes de veure que el teu amic s'ha entrebancat i ha caigut. Què li diries?"('Your friend just tripped and fell down. What would you say?').*

[1] Crystal, D. 1982. *Profiling linguistic disability*. London, UK: Edward Arnold.

[2] Shriberg, L. D., Kwiatkowski, J., Rasmussen, C. 1990. *The Prosody-Voice Screening Profile*. Tucson, AZ: Communication Skill Builders.

[3] Nowicki, S., Jr., Duke, M. P. 1994. Individual differences in the nonverbal communication of affect: The Diagnostic Analysis of Nonverbal Accuracy Scale. *Journal of Nonverbal Behavior* 18, 9–35.

[4] Peppé, S. & McCann, J. 2003. Assessing intonation and prosody in children with atypical language development: The PEPS-C test and the revised version. *Clin Linguist Phon* 17, 345-354.

[5] Klieve, S. A. 1998. *Perception of prosodic features by children with cochlear implants. Is it sufficient for understanding meaning differences in language*. Unpublished master's thesis. University of Melbourne, Australia.

[6] Lai, Z., Hughes, S., Shapiro, E. 1991. *Manual for the Minnesota Tests of Affective Processing (MNTAP).* University of Minnesota, MN.

[7] Prieto, P., Esteve-Gibert, N. (eds). 2018. *The development of prosody in first language acquisition*. Amsterdam: John Benjamins.

[8] Vanrell et al. 2018. The Discourse Completion Task in Romance prosody research: Status quo and outlook. In Feldhausen, I., Fliessbach, J. & Vanrell, M. (Eds.), *Methods in prosody: A Romance language perspective*. Berlin: Language Science Press, 191-229.

[9] Carrow-Woolfolk, E. 2017. *Comprehensive assessment of spoken language, Second Edition (CASL-2)*. Torrance, CA: Western Psychological Services.

[10] Wiig, E. H., Semel, E., & Secord, W. A. 2013. *Clinical Evaluation of Language Fundamentals–Fifth Edition (CELF-5)*. Bloomington, MN: NCS Pearson.

[11] Phelps-Terasaki, D., & Phelps-Gunn, T. 2007. *The Test of Pragmatic Language (2nd ed.)*. Austin, TX: Pro-Ed.

Audiovisual speech and the discrimination of a native fricative contrast: adult and infant data

Marta Ramon-Casas, Natasha Sanz Gauntlett, Ferran Pons and Laura Bosch
University of Barcelona

The study of phonetic discrimination in infancy has revealed perceptual changes resulting from exposure to the language in the environment, typically with enhanced discrimination for native contrasts and a gradual decline for non-native ones taking place during the second semester of life [1,2]. Some native contrasts, however, seem to challenge this developmental pattern as they appear to be initially difficult to discriminate and require longer exposure till infants reach the expected performance in discrimination tasks [3]. Fricative contrasts are among those that show a less uniform developmental pattern. Although not extensively explored, evidence from infant's discrimination studies (auditory modality) offers mixed results, especially regarding place of articulation discrimination for voiceless fricatives [4-7]. The present research focuses on a native fricative place of articulation contrast (/f/-/s/) and explores whether discrimination is enhanced when redundant audiovisual cues are provided.

Previous research in our lab failed to find evidence of discrimination in six-month-old infants for this native fricative place of articulation distinction, involving a sibilant and a non-sibilant voiceless pair, using CVC natural speech stimuli (fit vs. sit) produced by six different speakers and an auditory-only presentation of the test material. In a follow up experiment, a new group of infants was tested with a selection of the same material, based on acoustic measures of infant-directed speech properties, assessing whether this cue could enhance discrimination [8]. Again, no positive evidence could be gathered. An AV discrimination experiment was then designed. Beginning around six months of age infants' attention to a talker's mouth gradually increases with the detection of the perceptual equivalence between auditory and visual properties of speech and likely connected to speech processing [9]. Thus it was expected that AV cues could support discrimination of the target contrast. A total of N=33 six-month-old infants from Catalan and Spanish families (/f/ and /s/ are contrastive in both languages) were tested using the habituation-switch procedure. The AV material involved four different speakers producing tokens of either [sit] or [fit] stimuli (half of the sample was habituated to [f] tokens, the other half to [s] tokens). At test, attention to two same and two switch trials was measured, with longer attention to the switch trials indicating discrimination of the contrast. An experiment with adult participants was also planned. Adults (N=29) were undergraduate students from Catalan and Spanish backgrounds who were requested to identify the initial fricative consonant of [fit] and [sit] stimuli, also produced by four different speakers. The experiment involved the three modalities (visual, auditory and AV) for a total of 192 tokens presented in random order.

Infant data yielded no clear indication of discrimination (Fig. 1) with similar looking time measures to same and switch test trials ($F < 1$). Adults' accuracy scores (Fig. 2), as expected, revealed no difficulties in the task, but a highly significant effect of modality ($p < 0.0001$), with a lower performance in the visual-only modality, and a significant interaction of modality x consonant ($p = 0.02$), with a small advantage for [s] tokens in the AV condition. Contrary to our expectation, successful categorization and discrimination of this contrast could not be confirmed in six-month-old infants, even with AV speech material. Adult data confirmed that AV cues are useful in discrimination, although no clear advantage over the auditory condition was found due to the high level of performance in both conditions. A follow up study with older infants has been planned to establish the age at which discrimination is finally attained and to explore the role played by phonetic context ([fat]-[sat] stimuli) in the discrimination.

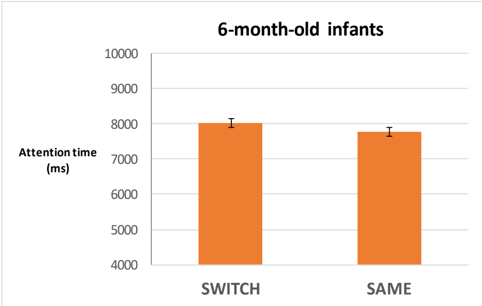


Figure 1. Mean attention time to switch and same test trials after habituation to [sit] or [fit] AV stimuli

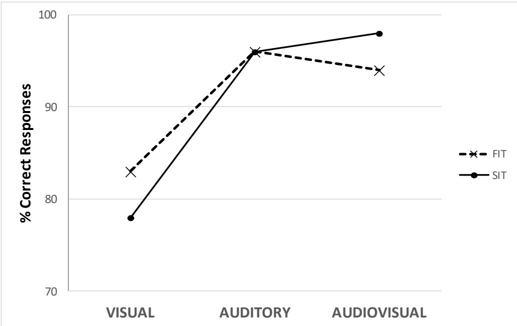


Figure 2. Accuracy in [s]-[f] identification by adult participants in three different presentation modalities.

[1] Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental science*, 9(2), F13-F21.

[2] Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant behavior and development*, 7(1), 49-63.

[3] Narayan, C. R., Werker, J. F., & Beddor, P. S. (2010). The interaction between acoustic salience and language experience in developmental speech perception: Evidence from nasal place discrimination. *Developmental Science*, *13*(3), 407-420.

[4] Eilers, R. E., Wilson, W. R., & Moore, J. M. (1977). Developmental changes in speech discrimination in infants. *Journal of Speech, Language, and Hearing Research*, *20*(4), 766-780.

[5] Tyler, M. D., Best, C. T., Goldstein, L. M., & Antoniou, M. (2014). Investigating the role of articulatory organs and perceptual assimilation in infants' discrimination of native and non-native fricative place contrasts. *Developmental psychobiology*, *56*(2), 210-227.

[6] Cristià, A., McGuire, G. L., Seidl, A., & Francis, A. L. (2011). Effects of the distribution of acoustic cues on infants' perception of sibilants. *Journal of phonetics*, *39*(3), 388-402.

[7] Beach, E. F., & Kitamura, C. (2011). Modified spectral tilt affects older, but not younger, infants' native-language fricative discrimination. *Journal of Speech, Language, and Hearing Research*, *54*(2), 658-667.

[8] Konjevod, M. (2015). *Can infant-directed speech facilitate the discrimination of a native fricative contrast in 6-month-old infants?* Unpublished masters' thesis. Cognitive Science and Language program (University of Barcelona).

[9] Lewkowicz, D. J., & Hansen-Tift, A. M. (2012). Infants deploy selective attention to the mouth of a talking face when learning speech. *Proceedings of the National Academy of Sciences*, *109*(5), 1431- 1436.

[ˈpɾa.tʊ], [ˈpa.tʊ], [ˈpla.tʊ], [ˈpᵉɾa.tʊ], [ˈpa.tɾʊ], [ˈpaɾ.tʊ]: Variable outputs for CCV syllables in the acquisition of Brazilian Portuguese

Andressa Toni (andressa.toni@usp.br)
University of São Paulo

This study explores the variable production patterns observed in the development course of CCV branching onsets (Consonant$_1$+Consonant$_2$+Vowel) in Brazilian Portuguese. Branching onsets are pointed as both articulatorily and phonologically complex for language acquisition, and although these syllables are fully developed only around 5 years old,[1] common words containing CCV may figure as targets in child speech even before 2 years old: *abre aqui* 'you open here'; *obrigado* 'thank you' (1;9 years old)[2]. Until CCV is fully acquired, branching onsets are often produced by repair strategies meant to modify the CCV structure, as in (1); or to modify the CCV segmental content, as in (2):

(1) ***gru**dei* 'I sticked': [**gu**ˈdej], [**guɾ**ˈdej], [**gu.ɾu**ˈdej], [**gu**.deˈ**ɾ**ej];

(2) ***gru**dado* 'It's sticked' [**glu**ˈda.dʊ]; ***cla**sse* 'classroom' [ˈ**kɾa**.sɪ]; ***tra**ve* 'goalpost' [ˈ**t**a.**vɾ**ɪ]

Both (1) and (2) repair patterns can co-occur with adult-like CCV productions in children's speech, either in longitudinal naturalistic data or in transversal experimental data.[2] Since these strategies are systematic within and across subjects and languages, Fikkert (2010)[3] raises two different accounts to analyze these variable patterns. In one view, children's variable outputs are due to unspecified stored lexical representations, which become more detailed as the child's grammar develops. In an alternative view, these repair strategies are reflecting adult-like stored lexical forms that are affected by articulatory constraints and by an underdeveloped phonological grammar that force correct inputs into the typical child repaired outputs. Considering these two possible accounts, the goal of the present study is to discuss whether children stored forms would have specified CCV clusters. Two clues were tested in a repetition task with 49 children between 2;4-5;10 years old:

(i) phonotactical clues: type and proportion of repair strategies for balanced CCV targets;

(ii) phonological clues: CCV palatalization. In Brazilian Portuguese, /t, d/ can be palatalized to [tʃ, dʒ] when in front of coronal high vowels [i, ɪ]. In child productions, the segmental context /t, d/ + [i, ɪ] can be generated when C$_2$ is deleted in CCV syllables like /tɾi, dɾi/. Applying the palatalization rule in these contexts would point to the adjacency of /t, d/ and [i, ɪ] in the child's stored word form, suggesting the phonological absence of C$_2$.

The corpus was organized into five groups based on the percentage of CCV adult-like productions, from G1 (0-5%) to G5 (76-100%). Results show that only subjects in G1 tended to modify the CCV structure as in (1); subjects from G2 to G5 tended to modify the segmental content of CCV as in (2) (cf. Table 1). G2 presented C$_2$ substitutions in both liquid directions in similar proportions (/l/→[ɾ]; /ɾ/→[l], cf. Table 2), suggesting that the segmental properties of the consonant in C$_2$ may be initially not fully specified. Palatalization was categorically applied by only one G1 child, while categorically blocked by most of the participants. However, 9 children from G1-G4 only partially blocked palatalization, which points to instabilities in CCV structural representation. These results point towards the unspecified nature of CCV syllables in child stored words, supporting Fikkert's account for an unspecified initial representation of the input. A pilot study testing CCV perception also points to underlying instabilities in both the identification of CCV/CV minimal pairs and the detection of CCV mispronunciations. In this pilot study, structural properties of CCV were discerned earlier than specific segmental properties of C$_2$ consonants, endorsing the observed in the production data (cf. (3)). Therefore, the present study supports the view that there are differences between the lexical representations of children and adults: the variability observed in child outputs can be due to the process of specification of the stored lexical representations. The specification process first addresses the structure of the syllables and then accounts for the segmental properties of the C$_2$ consonants in CCV branching onsets.

[1] LAMPRECHT, R. R (1993). A aquisição da fonologia do Português na faixa etária dos 2:9-5:5. Letras de Hoje, v.28, n.2.

[2] TONI, A (2017). Representação subjacente do ataque ramificado CCV na aquisição fonológica. Master’s dissertation.

[3] FIKKERT, P. (2010). Developing representations and the emergence of phonology: evidence from perception and production. In: Fougeron; Kühnert; D'Imperio (ed.), Laboratory Phonology 10: Variation, Phonetic Detail and Phonological Representation, p.227-258.

Table 1: Percentage of repair strategies for each production group

		G1 (0-5%)	**G2** (6-40%)	**G3** (41-60%)	**G4** (61- 75%)	**G5** (76-100%)	**Example** target > repaired
Structural modifying	*C_2 deletion*	70.59%	26.76%	5.8%	10.8%	6.3%	brilho ‘*sparkle’* [**bɾi**.ʎʊ] > [ˈ**bi**.ʎʊ]
	Epenthesis	0.36%	20%	30.46%	26.16%	12.71%	Pluto ‘*Pluto’* [ˈ**plu**.tʊ] > [**pʊ**ˈ**lu**.tʊ]
	Metathesis	0.73%	2.35%	1.72%	0.43%	0%	Pluto ‘*Pluto’* [ˈ**plu**.tʊ] > [ˈ**pul**.tʊ]
Content modifying	*C_2 substitution*	9.9%	42.35%	39.65%	22.36%	19.49%	bruxa ‘*witch’* [ˈ**bɾu**.ʃɐ] > [ˈ**blu**.ʃɐ]
	Transposition	0.73%	3.9%	3.45%	4.27%	0%	trave ‘*goalpost’* [ˈ**tɾa**.vɪ] > [ˈ**ta**.v**ɾ**ɪ]
	Reciprocal Movement	0%	0%	0%	0.43%	1.7%	trigo ‘*wheat’* [ˈ**tɾ**i.gʊ] > [ˈ**gɾ**i.**t**ʊ]
	Palatalization	31.25%	24.5%	33.33%	10.92%	0%	‘triste’ *sad* [**tɾi**s.tʃɪ] > [ˈ**tʃi**s.tʃɪ]

Table 2: Direction of the C2 substitutions for each production group

	G1 (0-5%)	**G2** (6-40%)	**G3** (41-60%)	**G4** (61- 75%)	**G5** (76-100%)	**Example** target > repaired
/l/→ [ɾ]	11.11%	47.23%	81.69%	64.15%	65.22%	‘classe’ *class* [ˈ**kla**.sɪ] > [ˈ**kɾa**.sɪ]
/l/→ glide	48.15%	4.64%	2.82%	0%	17.39%	‘Blico’ *Blico* [ˈ**bli**.kʊ] > [ˈ**b**^{w}**i**.kʊ]
/ɾ/→ [l]	25.93%	46.28%	5.63%	30.19%	17.39%	‘bravo’ *angry* [ˈ**bɾa**.vʊ] > [ˈ**bla**.vʊ]
/ɾ/→ glide	14.81%	1.85%	9.86%	5.66%	0%	‘Draco’ *Draco* [ˈ**dɾa**.kʊ] > [ˈ**d**^{j}**a**.kʊ]

(3) The pilot study tested the detection of minimal pairs and mispronunciations for CCV syllables turned into CV ([ˈpɾa.tʊ] > [ˈpa.tʊ]); CCV turned into CVC ([ˈpɾa.tʊ] > [ˈpaɾ.tʊ]); C/l/V turned into C[ɾ]V ([ˈpla.kɐ] > [ˈpɾa.kɐ]; and C/ɾ/V turned into C[l]V ([ˈpɾa.tʊ] > [ˈpla.tʊ]). Preliminary results show that children between 2 and 3 years old could not distinguish when CCV was mispronounced as CV, but could distinguish when it was mispronounced as CVC; children up to 4 years old could not distinguish when the liquids /l, ɾ/ were interchanged in CCV, but could distinguish them when in CV structures (to appear).

Why tune or text?
Explaining crosslinguistic variation in the resolution of tune-text conflicts

Marina Vigário, Marisa Cruz and Sónia Frota
University of Lisbon

Tonal material must be realized onto segments and patterns of temporal coordination are usually defined between tones and metrical or structural positions. However, segmental strings do not always contain enough voiced segmental material for the realization of complex tones or tone movements [1-7].

Languages, language varieties and dialects display an array of strategies to deal with emerging tune-text conflicts (see Table 1). If we exclude languages with variable alignment ([7]), strategies can be described as falling into two broad classes: (a) those that preserve the segmental material, at the expense of changes in tonal realization (compression, undershooting, tone truncation) – from hereon *Segmental Preservation Strategies* (SPS); and (b) those that preserve the tonal material, with impact on the segmentals (lengthening, vowel split, vowel epenthesis, blocking of vowel deletion or semivocalization) – from hereon *Tune Preservation Strategies* (TPS). In addition, there seems to be a general tendency for each language/variety to show more than one strategy to resolve the temporal conflict, but languages/varieties that select TPS tend not to make use of SPS, and vice-versa.

In this research we aim at contributing to the understanding of why languages show a preference for either TPS or SPS. On the basis of data from different languages and language varieties and dialects, we propose that the choice between strategies that preserve either the tune or the text springs from more general properties of the languages. Languages that organize their phonology predominantly around lower phonological levels care more for the segmentals, the tonal changes being tolerated and possibly subordinated to preserving the text; languages whose phonology is predominantly organized around higher, phrasal levels, by contrast, care more about the tune, segmental changes being more tolerated and possibly subordinated to preserving or enhancing the suprasegmentals.

The main empirical basis for this hypothesis is data previously analyzed from two varieties of Portuguese: Brazilian Portuguese (BP) and Standard European Portuguese (SEP), which implement strategies belonging to distinct groups, respectively SPS and TPS ([5-6]; Table 1 and Fig.1). In BP, many facts suggest that the phonology clusters predominantly around segments and lower prosodic domains: (i) systematic vowel epenthesis to fulfill well-formedness constraints at the syllable level; (ii) binary rhythmic stress; (iii) clipping phenomena at the foot level; (iv) nearly every word bears a pitch accent. In EP, productive phonological processes tend to involve higher domains and phenomena that span phrasal domains, including: (i) massive vowel deletion creating sequences of 6 and more consonants, triggered by prosodic reasons (context: unstressed position; consequence: enhanced word stress); (ii) unclear status of the syllable, due to unstressed vowels' deletion (resulting consonant sequences may remain unsyllabified); (iii) no evidence for the foot level - in normal speech no rhythmic stress, no foot-based clipping; (iv) lower prosodic levels do not bear pitch accent; (v) only the head of the IP is obligatory assigned a pitch accent, resulting in sparse tonal distribution and enhanced marking of IP head ([6, 8], a.o.).

An OT analysis is offered, innovatively considering the interaction of tonal and segmental constraints in lower and higher domains, expressing through a different ranking in both varieties the relevance of low domains and segmental preservation in BP, and of high domains and tone preservation in EP. Data coming from a meta-analysis of the literature will be presented supporting our hypothesis that the phonological profile as defined by the predominant types of phonological phenomena in the language predict the kind of strategies selected to deal with tune-text conflicts.

Table 1 – Tune-text accommodation strategies ([1-7], a.o.)

Accommodation strategies	Language/Language variety
Truncation (tonal targets are not realized)	Swedish; Northern Standard German; Palermo Italian; Friulian; Brazilian Portuguese – Atlantic Coast; European Portuguese – Braga; Leeds English; Belfast English; Moldavian Romanian; Catalan
Undershooting (tonal targets are partially realized)	Northern Standard German; Seoul Korean; Catalan
Compression (tonal movements are realized faster)	Swedish; Southern Standard British English; Northern Standard German; Catalan; Cambridge English; Newcastle English; Seoul Korean
Re-alignment (leftward shift of tonal targets)	Neapolitan Italian; Dutch; German
Lengthening (segments are lengthened)	Bari Italian; Standard European Portuguese; Fataluku
Split (1 vowel splits into 2 vowels)	Standard European Portuguese; European Portuguese – Funchal
Epenthesis (vowel insertion)	Standard European Portuguese; European Portuguese – ALE; European Portuguese – ALG; Bari Italian; Tashlhiyt Berber
Blocking of deletion/semivocalization (a process originating insufficient segmental space is blocked)	Standard European Portuguese; European Portuguese – ALG; European Portuguese – Funchal

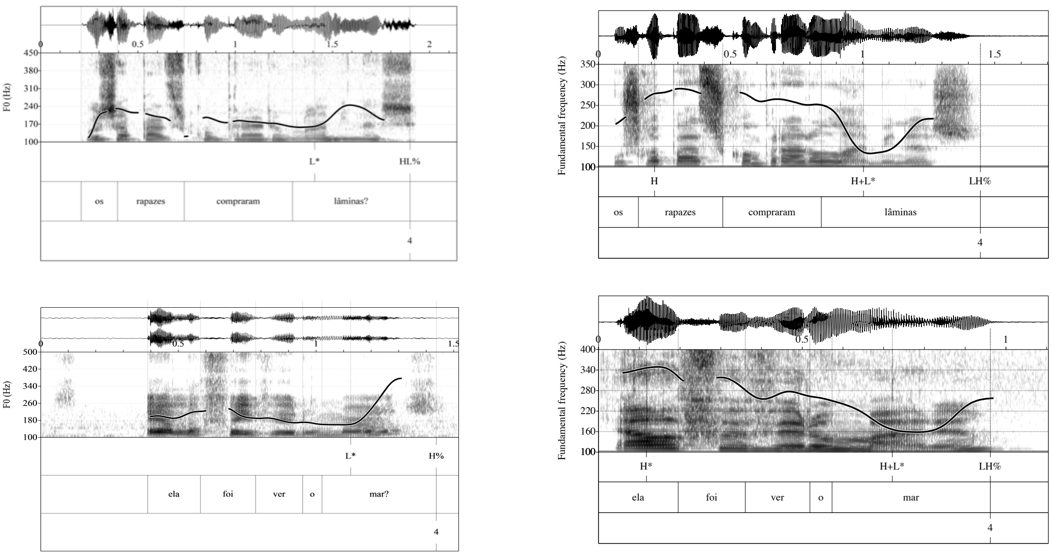


Figure 1. *Yes-no question contour in Southern BP (left) and SEP (right), without (top) and with (bottom) tune-text conflicts: truncation in BP and epenthesis in SEP (*['ma.ɾɨ]*).*

Selected references

[1] Grabe, E. 1998. Pitch accent realization in English and German. *JPhonetics*, 26(2): 129-143.

[2] Grabe, E., Post, B., Nolan, F., & Farrar, K. 2000. Pitch Accent Realization in Four Varieties of British English. *JPhonetics*, 28: 161-185.

[3] Frota, S. 2002. Tonal association and target alignment in European Portuguese nuclear falls. In Gussenhoven & Warner (eds.), *Laboratory Phonology 7*. Berlin: Mouton, 387-418.

[4] Grice, M., Savino, M., Caffo, A., & Roettger, T. B. 2015. The tune drives the text: schwa in consonant-final loan words in Italian. In *Proceedings of the 20th ICPhS, Glasgow*. Glasgow, UK.

[5] Frota, S., Cruz, M., Castelo, J., Barros, N., Crespo-Sendra, V., & Vigário, M. 2016. Tune or Text? Tune-text accommodation strategies in Portuguese. *Proc. of the Speech Prosody 2016*. Boston.

[6] Frota, S., & Prieto, P. 2015. *Intonation in Romance*. Oxford: Oxford University Press.

[7] Grice, M., Roettger, T.B., Ridouane, R. 2015. Tonal association in Tashlhiyt Berber. Evidence from polar questions and contrastive statements. *Phonology*, 32(2): 241-266.

[8] Wetzels, L., Menuzzi, S., & Costa, J. 2016. (eds.) *Handbook of Portuguese Linguistics*. Malden: Willey-Blackwell.

Oral session 10

Speech production

Intrasegmental Gestural Timing for American English /ɹ/ in Isolated and Connected Speech

Sarah Harper[1]
[1]*University of Southern California*

Recent research on intraspeaker variability in speech production has observed that many acoustic and articulatory attributes of speech vary systematically across different speech tasks, presumably as a consequence of greater temporal constraints on articulatory movement in faster, more casual speech [1]. One domain in which the effects of these temporal constraints on articulation are observed is in the degree of temporal overlap between discrete speech units (such as articulatory gestures), with the amount of overlap exhibited by adjacent gestures increasing at faster speech rates [e.g., 2]. However, as most existing research examining task and rate effects on gestural timing has focused on intersegmental timing relationships, the extent to which similar contextual effects are observed in intrasegmental gestural timing for multi-gesture consonants is unknown.

This study expands on previous work in both intergestural timing and contextual variability in speech production by examining the effect of speech task on the relative timing of two lingual gestures involved in the production of American English /ɹ/. American English /ɹ/ is widely characterized as involving the coordination of three supralaryngeal gestures, two of which are lingual: a palatal constriction made with either the tongue tip or body, and a pharyngeal constriction made with the tongue root [3]. The sequencing of these gestures has been shown to differ depending on the position of /ɹ/ in the word: the palatal gesture precedes the pharyngeal gesture in syllable-initial position and follows the pharyngeal gesture in syllable-final position. This positional difference in gestural sequencing has been shown to parallel observed differences in gestural magnitude for these constrictions across syllabic positions, leading to proposals that these magnitude asymmetries may explain the similarly asymmetric timing pattern [4]. As recent findings demonstrate systematic variation in the magnitude of both the palatal and pharyngeal gestures in /ɹ/ across speech tasks [5], the examination of speech task effects on timing enables a more thorough investigation of the relationship between gestural timing and magnitude, and enhances our understanding of the scope of gradience and variability in intrasegmental timing, an issue of importance in due to the role of articulatory variation in sound change [6].

Articulatory data for this study was taken from recordings of 40 speakers in the Wisconsin x-ray Microbeam (XRMB) database [7]. A total of 3,468 tokens of word-initial and word-final /ɹ/ were taken from three of the experimental tasks included in the corpus and separated into two conditions for analysis: (1) an **Isolated Speech** condition containing data from a Citation Word task (990 tokens) and (2) a **Connected Speech** condition containing data from Sentence and Prose Passage reading tasks (2,478 tokens). For each token of /ɹ/, maximum constriction time was measured for both the palatal and pharyngeal constrictions, with the gestural timing lag (**MLag**) calculated as the difference between the two maximum constriction times for a given token. The maximum palatal constriction degree and pharyngeal retraction were also measured to gauge the magnitude of each gesture.

The analysis of the MLag measurement indicates that although gestural sequencing remains constant between the Isolated and Connected Speech conditions, with the expected positional timing asymmetry observed in both conditions, the timing of these two gestures differs significantly across conditions. Specifically, less extreme timing differences are observed in the Connected Speech condition than in the Isolated Speech condition, mirroring the pattern of generally reduced gestural magnitude in the Connected Speech condition and reflecting an observed relationship between gestural magnitude and timing in this data.

[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] Browman, C.P., & Goldstein, L. 1984. Tiers in Articulatory Phonology, with Some Implications for Casual Speech.
[3] 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.
[4] Campbell, F., Gick, B., Wilson, I., & Vatikiotis-Bateson, E. 2010. Spatial and Temporal Properties of Gestures in North American English /R/. *Language and Speech*, 53, 49-69.
[5] Harper, S., Goldstein, L., & Narayanan, S. 2017. Stylistic Effects on the Acoustic and Articulatory Properties of English Rhotics. Poster presented at PaPE 2017, Cologne, Germany.
[6] Lawson, E., Stuart-Smith, J., and Scobbie, J.M. 2018. The role of gesture delay in coda /r/ weakening: An articulatory, auditory and acoustic study. *Journal of the Acoustical Society of America*, 143(3), 1646-1657.
[7] Westbury, J. 1994. *X-ray Microbeam Speech Production Database User's Handbook*. University of Wisconsin, Madison, WI.

Vertical larynx actions and larynx-oral timing in ejectives and implosives

Miran Oh, Dani Byrd, Louis Goldstein and Shrikanth S. Narayanan
University of Southern California

Glottalic airstream consonants such as ejectives and implosives are produced by initiating airflow in the supralaryngeal vocal tract by means of changing vertical larynx position [1, 2, 3]. For instance, ejectives and implosives involve raising or lowering of the larynx, respectively, which can consequently increase/decrease supralaryngeal air pressure. Such vertical movement of the larynx is of course not unique to non-pulmonic consonants, as voicing, pitch changes, swallowing, etc. also involve changes in the vertical larynx position. While correlations of vertical larynx movement with tone and intraoral pressure are known, little articulatory data exists regarding vertical larynx action and its timing with oral constriction formation in non-pulmonic consonants. For example, no clear distinction has been identified in the articulatory actions responsible for the phonological contrast between voiced implosives and voiced pulmonic stops, as both permit larynx lowering to decrease oral air pressure and maintain voicing through oral cavity expansion [3, 4, 5].

This study examines vertical larynx movement in the production of Hausa ejectives and implosives as well as in their pulmonic counterparts, using real-time Magnetic Resonance Imaging (MRI) data at 83 frames/sec of the midsagittal vocal tract obtained from a native female Hausa speaker. The test consonants include ejectives (/k’, k^{w}’, s’/) and their pulmonic counterparts (/k, k^{w}, s/), as well as implosives (/ɓ, ɗ/) and voiced stops (/b, d/). These target consonants are placed word-initially in bi-syllabic words with a LH tonal pattern, followed by a vowel /ɑ/. Data acquisition followed a real-time MRI protocol developed and extended in [6, 7], and region-of-interest image sequence analysis [8] and novel centroid tracking [9] were used to extract kinematic trajectories of the articulators for spatiotemporal study.

Findings on the magnitude of vertical larynx movement indicate that among ejectives, ejective fricatives (/s’/) have greater maximal raising of the larynx compared to ejective stops (/k’, k^{w}’/) (Figure 1: left). This is possibly due to fricatives requiring the maintenance of sufficient air flow for turbulence. A contrast between voiced implosives and voiced stops in their degree of larynx lowering is not observed (Figure 1: right). The results on the relative timing between vertical larynx movement and the coordinated oral constriction are presented in Figure 2. In both ejectives and implosives, the temporal lag from oral closure to larynx onset (Figure 2 left) is near zero (dotted lines indicate the median), suggesting that the vertical larynx movement is closely coordinated with the oral closure achievement. In contrast, in voiced stops larynx lowering starts much earlier. When the lag between oral closure *onset* and larynx onset lag is examined (Figure 2 right), lags in voiced stops are variable, but centered around zero, indicating that the oral constriction formation and the larynx lowering start approximately simultaneously. In contrast, the strong positive onset-to-onset lags seen in glottalic consonants indicate that vertical larynx movements are initiated much later for both glottalic consonants.

In conclusion, the current findings suggest that implosives and voiced stops contrast in the relative timing of vertical larynx movement with respect to their coordinated oral closures, rather than differing in the magnitude of their larynx lowering. The findings are in line with phonetic descriptions of ejectives as characterized by increased oral air pressure, whereas implosives are better explained with the absence of any pressure increase [2, 4], though there may be no significant pressure decrease. Significantly, the observed coordination between vertical larynx movement and the oral closure can be analyzed in-phase coupling for pulmonic stops, but anti-phase coupling for glottalic consonants. This extends the coupled oscillator model of Goldstein et al. [10] to capturing contrasts in airstream mechanism via the organization of gestures within a complex segment. [Supported by NIH]

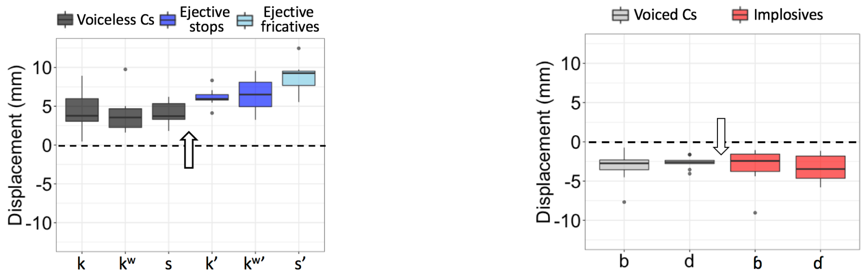


Figure 1. *Vertical larynx displacement*
(left: ejectives & voiceless consonants, right: implosives & voiced consonants)

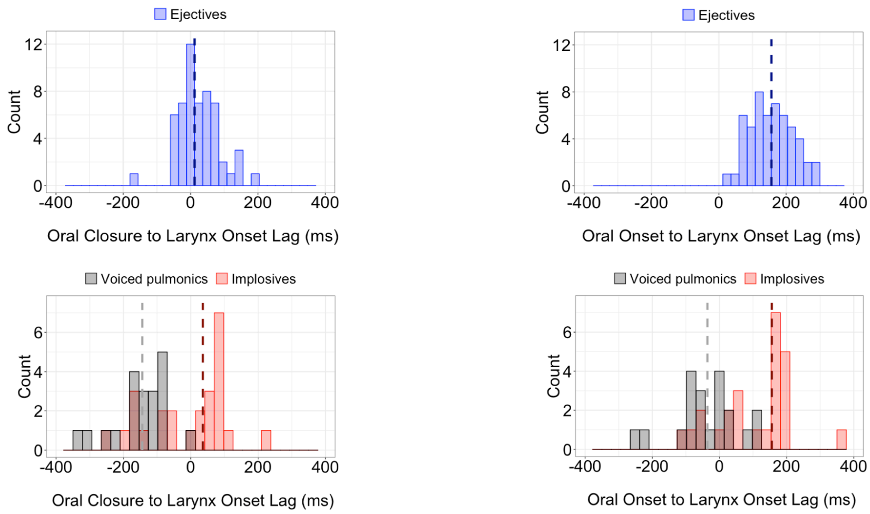


Figure 2. *Timing between oral gesture and vertical larynx movement*
(left: oral closure to larynx onset lag, right: oral onset to larynx onset lag)

[1] Catford, J. C. (1971). *Fundamental Problems in Phonetics.* Edinburgh: Edinburgh University Press.

[2] Greenberg, J. H. (1970). Some generalizations concerning glottalic consonants, especially implosives. *International Journal of American Linguistics*, *36*(2), 123-145.

[3] Ladefoged, P. (1968). *A Phonetic study of West African Languages: An Auditory-Instrumental Survey* (No. 1). Cambridge University Press.

[4] Clements, G. N., & Osu, S. (2002). Explosives, implosives and nonexplosives: the linguistic function of air pressure differences in stops. In C. Gussenhoven, & N. Warner (eds.), *Laboratory Phonology 7* (pp. 299-350). Berlin/New York: Mouton de Gruyter.

[5] Kingston, J. (1985). *The Phonetics and Phonology of the Timing of Oral and Glottal Events*. Doctoral dissertation. University of California, Berkeley.

[6] Narayanan, S. S., Nayak, K. S., Lee S., Sethy, A, Byrd, D. (2004). An approach to real-time magnetic resonance imaging for speech production. *Journal of the Acoustical Society of America, 115*(4), 1771-1776.

[7] Sajan Lingala, Yinghua Zhu, Yoon-Chul Kim, Asterios Toutios, Shrikanth Narayanan, Krishna Nayak. (2017). A fast and flexible MRI system for the study of dynamic vocal tract shaping. *Magnetic Resonance in Medicine*. 77(1), 112-125.

[8] Lammert, A., Ramanarayanan, V., Proctor, M., & Narayanan, S. (2013). Vocal tract cross-distance estimation from real-time MRI using region-of-interest analysis. In *INTERSPEECH* (pp. 959-962). Lyon, France.

[9] Oh, M., & Lee, Y. (2018). ACT: An Automatic Centroid Tracking tool for analyzing vocal tract actions in real-time magnetic resonance imaging speech production data, *The Journal of the Acoustical Society of America, 114*(4), EL290-EL296.

[10] Goldstein, L., Nam, H., Saltzman, E., & Chitoran, I. (2009). Coupled oscillator planning model of speech timing and syllable structure. *Frontiers in phonetics and speech science*, 239-250.

Network analytics reveals patterns in many-to-one articulatory-to-acoustic strategies

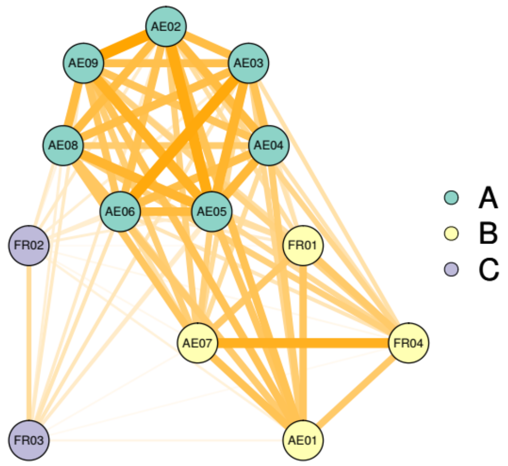
Christopher Carignan
Institute of Phonetics and Speech Processing - LMU Munich

With the many degrees of freedom involved in the articulation of speech sounds, any given acoustic dimension is rarely (if ever) modified by a single articulatory dimension. Rather, the acoustic consequences of speech are often influenced by multiple articulatory dimensions in a many-to-one relationship. A considerably ill-understood relationship of this sort involves the realization of F1 frequency in the production of vowel nasalization. Lowering the velum in order to couple the nasal cavity to the oro-pharyngeal cavity has been shown to modify the frequency of F1 in a manner that is independent of changes in tongue configuration [1]. However, both phonetic and phonological vowel nasality has been observed also to involve changes in tongue configuration [2-4] and increased breathy voicing [5,6], articulations which may modify F1 frequency in their own independent ways [7,8].

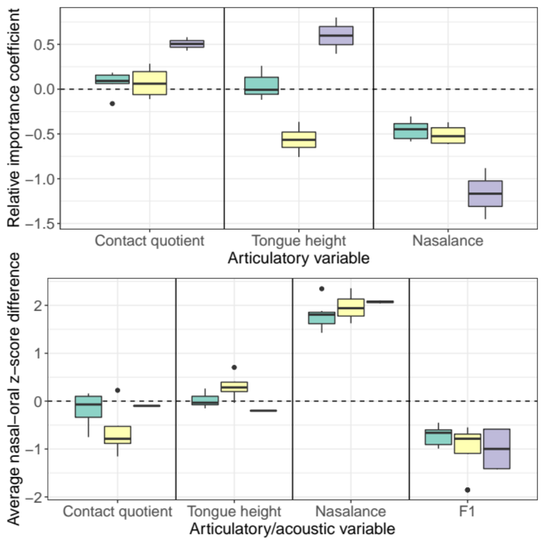
Articulatory data (nasalance, ultrasound, EGG) and acoustic data were co-registered to quantify the production of nasalization, tongue height, breathy voicing, and F1 frequency in the distinction of three nasal-oral vowel contrasts in French (FR). These data were collected first from four native FR speakers and, subsequently, from nine naïve Australian English (AE) listeners who imitated the FR productions [9]. For each of the 13 speakers, articulatory measurements were mapped to F1 measurements for each of the nasal-oral vowel pairs using relative importance analysis (RIA), and the RIA coefficients were used to create similarity scores among all of the speakers. These similarity scores were then used to build network models for each vowel pair, and the spinglass algorithm [10] was employed to identify communities/groups of shared articulatory-to-acoustic strategies within each network.

An example of the three-group network for the vowel pair /ɑ̃/-/a/ is shown in Figure 1. The group affiliations for the individual speakers are shown in subfigure (a). Although the overall variable distinctions are similar for the three groups (subfigure (b), bottom panel), the network model is able to identify differences in how these sub-groups of speakers use the three articulators to reach a similar acoustic output (subfigure (b), top panel): using only degree of nasalization (group A), using degree of nasalization and tongue height (group B), or using degree of nasalization, tongue height, and degree of breathiness (group C). In some cases, these strategies are language-dependent (groups A and C), while in other cases similar strategies are used by both FR and AE speakers (group B). The variety of strategies highlights the multi-dimensional nature of vowel nasality, rather than the uni-dimensional assumption of “nasal” vowels as merely oral vowels produced with a lowered velum [11]. Moreover, this multi-dimensionality is not necessarily limited to a native speaker’s phonological system, but it can be transmitted to listener-turned-speakers, which has important implications for models of listener-based sound change that involve the imperfect transmission of articulatory cues between speakers and hearers [12].

This study not only emphasizes the importance of recognizing the many-to-one nature of speech, but it also highlights the necessity of using both acoustic and articulatory data in research on speech production. In the example shown here, community distinctions are not observed in the measurements of the individual articulations, but are only revealed in the articulatory-to-acoustic strategies used by the speakers. Knowledge of these articulatory-to-acoustic relationships could not have been accurately predicted from the articulatory data alone, nor from the acoustic data alone. Research that seeks to gain a more comprehensive understanding of the complex nature of speech production should necessarily involve elements of speech articulation, speech acoustics, and the relationships between the two.



(a) 3-group network for /ɑ̃/-/a/



(b) RIA coefficients and variables

Figure 1. *Spinglass community network results for the /ɑ̃/-/a/ distinction (three groups). The network for the groups is displayed on the left sub-figure (a). The connection strength between any two participant nodes is represented by both line thickness and opacity. Values for each network group are shown in the right sub-figure (b) for the relative importance analysis coefficients (top panel) and the nasal-oral variable distinctions (bottom panel).*

[1] Carignan, C. (2018). Using ultrasound and nasalance to separate oral and nasal contributions to formant frequencies of nasalized vowels. *Journal of the Acoustical Society of America*, *143*(5):2588–2601.

[2] Barlaz, M., Shosted, R., Fu, M., and Sutton, B. (2018). Oropharygneal articulation of phonemic and phonetic nasalization in Brazilian Portuguese. *Journal of Phonetics*, *71*:81–97.

[3] Carignan, C., Shosted, R., Shih, C., and Rong, P. (2011). Compensatory articulation in American English nasalized vowels. *Journal of Phonetics*, *39*:668–682.

[4] Zerling, J. P. (1984). Phénomènes de nasalité et de nasalization vocaliques: Étude cinéradiographique pour deux locuteurs. *Travaux de l'Institut de Phonétique de Strasbourg*, *16*:241–266.

[5] Garellek, M., Ritchart, A., and Kuang, J. (2016). Breathy voice during nasality: A cross-linguistic study. *Journal of Phonetics*, *59*:110–121.

[6] Carignan, C. (2017). Covariation of nasalization, tongue height, and breathiness in the realization of F1 of Southern French nasal vowels. *Journal of Phonetics*, *63*:87–105.

[7] Stevens, K. N. (2000). *Acoustic phonetics*. MIT Press, Cambridge, MA.

[8] Lotto, A. J., Holt, L. L., and Kluender, K. R. (1997). Effect of voice quality on perceived height of English vowels. *Phonetica*, *54*:76–93.

[9] Carignan, C. (2018). Using naïve listener imitations of native speaker productions to investigate mechanisms of listener-based sound change. *Laboratory Phonology*, 9(1):1–31.

[10] Csardi, G. and Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal*, Complex Systems:1695.

[11] Shosted, R. (2015). Nasal vowels are not [+nasal] oral vowels. In Smith, J. and Ihsane, T., editors, *Proceedings of the 42nd Linguistic Symposium on Romance Languages (2012)*, Amsterdam. John Benjamins.

[12] Beddor, P. S. (2009). A coarticulatory path to sound change. *Language*, *85*(4):785–821.