

LMU

LUDWIG-  
MAXIMILIANS-  
UNIVERSITÄT  
MÜNCHEN

# Abstraction, Diversity, and Speech Dynamics

3-5 May 2017

Herrsching am Ammersee



CAS  
Research Focus

# Imprint

## Scientific Organizers:

Prof. Dr. Jonathan Harrington, PD Dr. Marianne Pouplier, PD Dr. Eva Reinisch  
(Institute of Phonetics and Speech Processing, LMU Munich, Germany)

Academic Coordinator: Dr. Lena Bouman (CAS<sup>LMU</sup>)

Institutional Responsibility: Center for Advanced Studies (CAS<sup>LMU</sup>)

Editor: PD Dr. Marianne Pouplier



Center for Advanced Studies  
Ludwig-Maximilians-Universität München  
Seestraße 13  
80802 München  
info@cas.lmu.de  
www.cas.uni-muenchen.de

Preface	2
Program	4
Abstracts Talks	8
Abstracts Poster Session 1	32
Abstracts Poster Session 2	57
Author Index	79
Program at a Glance	80

Deutsche  
Forschungsgemeinschaft  
**DFG**



Welcome to the workshop on **Abstraction, Diversity, and Speech Dynamics** hosted by the Institute of Phonetics and Speech Processing of the Ludwig Maximilians University Munich. We are pleased to welcome you all to Herrsching am Ammersee.

This workshop is held as part of the Research Focus *How Words Emerge and Dissolve* of the Centre for Advanced Studies of the LMU. The Research Focus is concerned with how the sounds of speech are acquired by infants and young children in first language acquisition, how such acquisition stabilizes in healthy individuals, and how such patterns may dissolve following the onset of brain lesions. An issue deeply intertwined with these questions is the relationship between continuous signal dynamics and the parallel transmission of numerous types of signs or categories, both linguistic (e.g. words and their constituent consonants and vowels) and indexical (social class, regional affiliation, gender etc.). While it has long become clear that the linguistic and social as well as the cognitive and physical aspects of speaking are tightly intertwined, quite how these multiple layers of semiotic and signal aspects of speech are connected and how those connections may be manifested differently in the world's languages and cultures remains poorly understood. The aim of the conference is to advance the discussion on these issues by bringing together scientists from various disciplines engaged in research on areas such as memory and its relationship to abstraction, feedback and feedforward control systems, and modelling the association between discrete categories and continuous speech dynamics. It is only with a deeper understanding of the semiotic-signal association that breakthroughs can be achieved in understanding how the sounds of language are acquired, in how normal and disordered mechanisms of speech are related, and in the way that social and linguistic information interact and are transmitted in speech communication. We are confident that the broad range of scientific work presented at this workshop provides a wonderful opportunity to open new vistas on these complex issues.

We would like to thank our colleagues at the IPS who have helped in many ways to make this workshop possible, in particular Lia Saki Bučar-Shigemori, Tabea Pape, Ulrike Vallender-Kalus and Katrin Wolfswinkler.

We gratefully acknowledge the support of the Centre for Advanced Studies of the LMU, the European Research Council (295573-SCATS; 283349-SCSPL), the German Research Council (PO 1269/4-1; RE 3047/1-1), and the Association of Laboratory Phonology which helped us publicize the workshop.

With best wishes for an enjoyable and productive meeting

*Jonathan Harrington, Marianne Pouplier, Eva Reinisch*

# Program

## Wednesday – May 3 2017

09:20 – 09:30

**Welcome Address**

**Session 1:**

*Chair Matthias Sjerps*

09:30 – 10:15

**James McQueen:**

Speech perception: as abstract as it needs to be

10:15 – 10:35

**Molly Babel, Brianne Senior, Karina Wong:**

A first attempt at modeling social preferences in perceptual learning

10:35 – 10:55

**Alessandro Vietti, Lorenzo Spreafico, Vincenzo Galatà, Constantijn Kaland:**

Mapping language experience onto phonetic categories

10:55 – 11:15

*Coffee Break*

**Session 2:**

*Chair Ander Egurtzegi*

11:15 – 12:00

**Janet Pierrehumbert:**

Learning socio-indexical features of words

12:00 – 12:20

**Laura Spinu:**

Survival and neutralization of a rare cross-linguistic contrast: the case of Romanian palatalized postalveolars

12:20 – 02:00

*Lunch Break*

**Session 3:**

*Chair Melissa Redford*

02:00 – 02:45

**Jan Edwards:**

When ears collide: mismatch and phonological development

02:45 – 03:05

**Núria Esteve-Gibert, Hélène Loevenbruck, Marion Dohen, Mariapaola D'Imperio:**

The use of prosody and gestures for the production of contrastive focus in French-speaking 4 and 5 year olds

03:05 – 03:30

*Coffee Break*

03:30 – 05:00

**Poster Session 1**

**Session 4:**

*Chair Cécile Fougeron*

05:00 – 05:20

**Wolfram Ziegler, Hanna Jakob, Katharina Lehner, Ingrid Aichert:**

A "Phonological Mind" in our brains? Evidence from patients with brain lesions

05:20 – 05:40

**Doris Mücke, Anna Hermes, Tabea Thies:**

Discrete categories and continuous speech dynamics in typical and atypical speech: How much variability is tolerated by the phonological system?

05:40

**Welcome Reception**

06:30

*Dinner*

## Thursday – May 4 2017

**Session 5:**

*Chair Molly Babel*

08:30 – 09:15

**Esther Janse:**

Spoken language processing across the adult life span

09:15 – 09:35

**Merel Maslowski, Antje S. Meyer, Hans Rutger Bosker:**

When slow speech sounds fast: How the speech rate of one talker influences perception of another talker

09:35 – 09:55	<b>Matthias Sjerps, Caicai Zhang, Gang Peng:</b> Unequal effects in the normalization of lexical tone and the normalization of vowel quality
09:55 – 10:15	<b>Helena Levy, Lars Konieczny, Adriana Hanulíková:</b> Long-term effects of accent exposure on perception and production
10:15 – 10:30	<i>Coffee Break</i>
10:30 – 12:30	<b>Poster Session 2</b>
12:30 – 02:00	<i>Lunch Break</i>
<b>Session 6:</b>	<i>Chair Doris Mücke</i>
02:00 – 02:45	<b>Pascal Perrier:</b> From abstract phonemes to speech movements: the role of orofacial biomechanics and multisensory motor goals
02:45 – 03:30	<b>Adamantios Gafos:</b> Nearly-decomposable hierarchies, symbolic grammars and pervasiveness of dynamics
03:30 – 04:00	<i>Coffee Break</i>
<b>Session 7:</b>	<i>Chair Philip Hoole</i>
04:00 – 04:45	<b>Douglas Shiller:</b> Auditory adaptation effects in speech motor control
04:45 – 05:30	<b>Caroline Niziolek:</b> How do communicative goals affect continuous speech dynamics?
05:30 – 05:50	<b>Mareike Flögel, Christian Kell:</b> Adaptation to dichotically presented spectral and temporal real-time perturbations of auditory speech feedback

06:30 *Dinner*

## Friday – May 5 2017

<b>Session 8:</b>	<i>Chair Simone Falk</i>
08:30 – 09:15	<b>Jennifer Cole:</b> Individual differences and attentional effects on cue weighting for prosody perception
09:15 – 10:00	<b>Ann Bradlow:</b> Language-independent talker-specificity in bilingual speech production
10:00 – 10:30	<i>Coffee Break</i>
<b>Session 9:</b>	<i>Chair Wolfram Ziegler</i>
10:30 – 11:15	<b>Matthew Goldrick:</b> Dynamics of lexical access and speech articulation
11:15 – 11:35	<b>Martin Corley, Eleanor Drake, Sara Knight:</b> Using ultrasound speech imaging to investigate within- and cross-language interference in word production
11:35 – 11:55	<b>Melissa Redford, Sara Pacchiarotti:</b> Temporal information marks person in a Northern Lazio Italian variety
11:55 – 12:15	<b>Farewell, Closing remarks</b>
12:30	<i>Optional Lunch (if booked with registration)</i>

## Speech perception: as abstract as it needs to be

**James McQueen**

Radboud University, The Netherlands

*j.mcqueen@donders.ru.nl*

To understand speech, listeners have to abstract away from the phonetic diversity which talkers confront them with. But they also need to store knowledge about how individual talkers and groups of talkers speak, so that they can understand those talkers better in the future. In this talk, I will briefly summarize evidence that these two processes are supported by complementary memory systems, one which stores abstract representations of speech and one which stores episodic memories that include talker and situational details. I will then turn to a more specific question: what linguistic abstractions do listeners use? Findings will be presented from recent speech learning experiments using lexically-guided retuning and selective adaptation paradigms. These studies show that listeners use context-dependent allophonic units, rather than context-independent phonemic units, in speech recognition. I will argue that this is because tuning in to speech at the level of the allophone helps the listener understand speech, while tuning in at the level of the phoneme does not. More generally, the degree of abstraction that is required in speech recognition, about phonological content and about talkers, is what is required for listeners to be able to cope with the diversity inherent in the speech signal.

## A first attempt at modeling social preferences in perceptual learning

Molly Babel, Brianne Senior, Karina Wong

Department of Linguistics, University of British Columbia, Canada

[molly.babel@ubc.ca](mailto:molly.babel@ubc.ca)

Humans possess an impressively robust and reliable speech recognition system. In the face of immense variability, listeners (generally) comprehend spoken language with ease. A trademark feature of the human speech recognition system is that it is stable, yet flexible. What affords us with this talent? A key mechanism claimed to provide this combination of flexibility and stability is *perceptual learning*: a change as a result of sensory exposure to a deviant or unexpected signal. Simply, what was previously not recognized as a member of a particular sound category or as an interpretable pronunciation of a particular word is now categorized or recognized as such (e.g., Norris et al., 2003). Here we ask whether social preferences impact perceptual learning. Listeners do not weight all incoming phonetic information equivalently (Johnson, 1997), with familiar accents, for example, benefiting from improved encoding (e.g., Clopper et al., 2016) and claims that listeners attend less to dispreferred accents (Lippi-Green, 1997). Listeners in our study were presented with a passage from Pinocchio where the back vowels were participating in an F1 raising chain shift (Weatherholtz, 2015). An actor-phonetician produced this story in a pleasant style (with shifted vowels), an unpleasant style (with shifted vowels), and control condition (with unshifted vowels). Perceptual learning was assessed with lexical decision endorsement (n=60, 20 per condition) and cross-modal priming (n=60, 20 per condition). The results of this study are a first step in understanding whether the mechanisms of perceptual learning may be affected by social preferences.

### References

- Clopper, C. G., Tamati, T. N., & Pierrehumbert, J. B. (2016). Variation in the strength of lexical encoding across dialects. *Journal of Phonetics*, 58, 87-103.
- Johnson, K. (1997). Speech perception without speaker normalization. *Talker variability in speech processing*, ed. by Keith Johnson and John W. Mullennix, 145–65. San Diego: Academic Press.
- Lippi-Green, R. (1997). *English with an accent: Language, ideology, and discrimination in the United States*. Psychology Press.
- Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, 47 (2), 204-238.
- Weatherholtz, K. (2015). *Perceptual learning of systemic cross-category vowel variation* (Doctoral dissertation, The Ohio State University).

## Mapping language experience onto phonetic categories

Alessandro Vietti<sup>1</sup>, Lorenzo Spreafico<sup>1</sup>, Vincenzo Galatà<sup>1</sup>, Constantijn Kaland<sup>2</sup>

<sup>1</sup>Alpine Laboratory of Phonetic Sciences Free University of Bozen-Bolzano, Italy;

<sup>2</sup>University of Cologne, Germany

[alessandro.vietti@unibz.it](mailto:alessandro.vietti@unibz.it)

Sociophonetic perspectives on linguistic variability (Docherty & Foulkes 2014; Harrington 2006) assume that adult speakers constantly update their phonetic representations as they integrate new acoustic, articulatory, and socio-indexical information.

In our proposal, we investigate the relation between the linguistic experience and the variability of sociophonetic categories in Italian-Tyrolean bilingual speakers. More precisely, we explore how social network properties (Carrington 2005) affect the production of rhotics by focusing on the duration of the constriction phase of /r/ in Tyrolean (German Dialect; Galatà et al. 2016).

We hypothesize that variability in language use and network density affect the duration of /r/s in Tyrolean (Sancier & Fowler 1997). To test this hypothesis we draw on a dataset coming from a large-scale experiment: we consider egocentric network data collected via electronic questionnaires as well as acoustic production data elicited using a list of Tyrolean words specifically designed to contain all possible contexts in which rhotics can occur. We discuss results from a dataset of 2248 tokens as produced by 14 subjects: 8 sequential bilinguals (Tyrolean-dominant), 6 simultaneous bilinguals; gender: 11 F, 3 M; age: mean = 26.5, sd = 5.31). We treated constriction duration as dependent variable and variability of language choice (measured as entropy of choice) and network density as main predictors. Results show that both variability of language choice and network density exert a significant influence on the duration of the constriction phase of /r/ in Tyrolean.

### References

- Carrington P., Scott J., Wasserman S. (eds.) 2005 *Models and Methods in Social Network Analysis*. Cambridge: CUP.
- Docherty G., Foulkes P. 2014 An evaluation of usage-based approaches to the modelling of sociophonetic variability. *Lingua*, 142 (04): 42-56.
- Galatà V. et al. 2016 An acoustic analysis of /r/ in Tyrolean, InterSpeech 2016. Harrington J. 2006 An acoustic analysis of 'happy-tensing' in the Queen's Christmas broadcasts. *Journal of Phonetics* 34. 439-457.
- Sancier M., Fowler C. 1997 Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics*, 25: 421-436.

## Learning socio-indexical features of words

**Janet Pierrehumbert**

Oxford University e-Research Centre, UK

*janet.pierrehumbert@oerc.ox.ac.uk*

People who speak the same language differ in what words they use and also in their pronunciation patterns. Some of these differences become conventionalised as socio-indexical, permitting listeners to make social inferences about the speaker while also grasping the content of what the speaker is saying. This situation points to cognitive processes whereby people can learn social associations for words and pronunciation patterns, and propensities to imitate others engender groups sharing linguistic behaviours. But the propensities to imitate others must be limited, as otherwise the lexical choices and speech patterns would become completely homogenised within any given speech community. In this talk, I will review a series of experimental studies about the power and limitations of socio-indexical learning. These studies indicate that:

- Indexical associations for words, morphological patterns, and allophonic patterns are all quite learnable.
- Categorical patterns can be learned with a remarkably small number of trials. Learning such patterns is much faster than learning gradient phonetic patterns.
- Only a subset of the available statistical associations are learned. Contextual relevance shapes learning by affecting what examples are remembered and how generalisations are formed.
- “Contextual relevance” derives both from widely shared assumptions about how language works, and from individual social identities.

## Survival and neutralization of a rare cross-linguistic contrast: The case of Romanian palatalized postalveolars

**Laura Spinu**

University of Toronto, Canada

*laura.spinu@utoronto.ca*

The contrast between plain and secondarily palatalized (SP) postalveolar fricatives is rare cross-linguistically [2], possibly due to low salience resulting from gestural timing [6]. This contrast is present in Romanian as a morphophonological number marker [4], despite native speakers not being perceptually sensitive to it [3]. The results of a 1961 study [5] suggest this situation has persisted for at least 50 years. This is somewhat unexpected, as perceptually fragile contrasts tend to be enhanced or neutralized [1]. We report the results of (a) a production experiment (n=31) and (b) a perception experiment (n=30), in which the stimuli included additional morphological cues to the presence of SP. We hypothesized that the additional morphological information might yield stronger perceptual effects, especially if conflicting with the information conveyed by SP. We found that, while 87% of the speakers produce significant differences between plain and palatalized postalveolar fricatives, listeners' sensitivity to the contrast is low. Fewer speakers produced the contrast compared to 1961, signaling neutralization in progress. Neutralization was stronger in males than females. Aside from documenting this phenomenon, an explanation is sought for its longevity and it is proposed that grammatical restructuring [2] offers the best account for the observed facts. These findings add further evidence to the claim that there is no 1-to-1 correspondence between the phonetic factors triggering neutralization and actual neutralization patterns attested in languages, and that an explanation for phonotactic patterns should be sought in the interaction of phonetic and sociolinguistic factors with the phonological grammar.

### References

- [1] Hayes, B. and Steriade, D. 2004. Introduction: The phonetic bases of phonological markedness. In Hayes, B., Kirchner, R. and Steriade, D. (eds.), *Phonetically-based phonology*. CUP.
- [2] Kochetov, A. 2002. *Production, Perception, and Emergent Phonotactic Patterns: A Case of Contrastive Palatalization*. New York: Routledge.
- [3] Spinu, L., Vogel, I., and Bunnell, T. (2012). Palatalization in Romanian: Acoustic properties and perception. *Journal of Phonetics* 40(1), 54–66.
- [4] Spinu, L. and Lilley, J. 2016. A comparison of cepstral coefficients and spectral moments in the classification of Romanian fricatives. *Journal of Phonetics* 57, 40–58.
- [5] Şuteu, V. 1961. Observații asupra pronunțării limbii române. *Studii și cercetări lingvistice* 12(3), 293–304.
- [6] Zsiga, E. 2000. Phonetic alignment constraints: Consonant overlap and palatalization in English and Russian. *Journal of Phonetics* 28, 69–102.

## When ears collide: mismatch and phonological development

Jan Edwards

University of Maryland, USA

[edwards@umd.edu](mailto:edwards@umd.edu)

Phonological development is the process by which children impose discrete categories on continuous speech dynamics, to learn to produce and perceive thousands of words over the first half dozen years of life. While input is crucially important in acquisition, it is not always the case that the linguistic experience and expectations of the speaker providing the input align with the linguistic experience of the language learner. In this talk, I will provide three very different examples of this phenomenon. The first focuses on the acquisition of the “same” L1 phonological contrast across two languages. In this case, differences in order of acquisition of phonemes and error patterns in the two languages can be explained by the differences in adult native-speaker judgments across the languages. The second example focuses on the acquisition of phoneme contrasts by children with cochlear implants. An analysis of error patterns produced by children with cochlear implants reveals an atypical pattern of development that is highly influenced by the spectral degradation of the input provided by the cochlear implant. Finally, the third example focuses on the impact of dialect mismatch on phonological development after literacy acquisition. Children who speak a “non-mainstream” dialect (such as African American English) at home experience dialect mismatch when they go to school where the Academic Classroom dialect is always aligned instead to a “mainstream” dialect. In order to learn to read and to acquire words from text as well as from speech, children who speak a non-mainstream dialect must learn to map from the phonological categories of their home dialect to those of the other dialect.

[Work supported by NIH]

## The use of prosody and gestures for the production of contrastive focus in French-speaking 4 and 5 year olds

Núria Esteve-Gibert<sup>1</sup>, Hélène Loevenbruck<sup>2</sup>, Marion Dohen<sup>3</sup>, Mariapaola D’Imperio<sup>1</sup>

<sup>1</sup>Aix Marseille Université, CNRS, LPL, Aix-en-Provence, France, <sup>2</sup>Univ. Grenoble Alpes, CNRS, LPNC UMR 2105, Grenoble, France, <sup>3</sup>Univ. Grenoble Alpes, CNRS, GIPSA-lab, Grenoble, France

[nuria.esteve-gibert@blri.fr](mailto:nuria.esteve-gibert@blri.fr)

This study examines prosodic and gestural correlates of contrastive focus in French-speaking children producing imperative sentences (e.g. ‘Take the ORANGE dress [not the blue dress]’). Previous research suggests that children do not seem to use adult-like prosodic cues to highlight elements in the discourse before the age of 8-10 years [1], although they do use phrasing strategies [2] and rhythmic co-speech gestures during pre-school ages [3]. We hypothesize that young children might produce contrastive focus but through body gestures and phrasing, before they master the complete set of adult-like pitch patterns. Forty French-speaking 4 and 5 year olds participated in a game that elicited spontaneous production of sentences in 3 conditions (no-focus; contrastive focus; corrective focus) and 2 target focus positions (on the noun; on the adjective). Children were audio-visually recorded and assessed for linguistic and hearing abilities. We analyzed prosodic (pauses; F0max, syllable duration, intonation) and gestural cues (gesture type; gesture-speech alignment). Preliminary findings reveal that 4-year-old children did not use adult-like French prosodic cues like initial rise, lengthening or de-accentuation [4]; instead, they seemed to use gestures and pauses, especially in the corrective focus condition. In contrast, 5-year-old children seemed to use gesture cues more systematically and were better at prosodic cues. These results will reveal the development of discourse prominence in young children (which might be observed early if information highlighting is seen as a multimodal phenomenon), and will help understanding the dynamic emergence of prosodic and gestural abilities for meaning purposes in children.

### References

- [1] K. Ito, “Children’s pragmatic use of prosodic prominence,” in *Pragmatic Development in First Language Acquisition*, 2014.
- [2] A. S. H. Romoren and A. Chen, “Quiet is the New Loud: Pausing and Focus in Child and Adult Dutch,” *Lang. Speech*, vol. 58, pp. 8–23, 2015.
- [3] M. Mathew, I. Yuen, S. Shattuck-Hufnagel, A. Ren, and K. Demuth, “The development of prosodic gestures during parent-child discourse interactions,” in *Laboratory Phonology*, 2014.
- [4] M. Dohen and H. Loevenbruck, “Pre-focal Rephrasing, Focal Enhancement and Post-focal Deaccentuation in French,” *8th International Conference on Spoken Language Processing*, 2004, pp. 2–5.

## A “Phonological Mind” in our brains? Evidence from patients with brain lesions

**Wolfram Ziegler, Hanna Jakob, Katharina Lehner, Ingrid Aichert**

EKN, Institute of Phonetics and Speech Processing, Ludwig Maximilians University, München, Germany

[wolfram.ziegler@ekn-muenchen.de](mailto:wolfram.ziegler@ekn-muenchen.de)

According to a theory proposed by Iris Berent, the human mind is equipped with a universal stock of “core phonological knowledge”, which is considered purely algebraic and autonomous from the perception/production channels of linguistic communication (Berent, 2013). As a part of the human genetic endowment, this knowledge is mediated by a dedicated neural network in the brain (Berent et al., 2014).

The present study was based on the assumption that sound production impairments resulting from lesions to this network provide insight into the architecture of the “Phonological Mind”. We examined (i) whether phonological impairment destroys the alleged “core phonological knowledge” postulated by Berent, and (ii) to what extent phonological impairment conforms to a model based on principles of intergestural coordination (Tilsen, 2016; Ziegler & Aichert, 2015). Fifteen patients with phonological impairment after left hemisphere infarctions were administered a word repetition task using 32 German nouns with varying phonological structures (1 – 4 syllables, simple vs. complex syllables, varying stress patterns). Their sound production errors were assessed by phonetic transcription. In a first analysis, the transcripts were examined for violations of core phonological regularities and of markedness constraints. In a second step, latent trait analyses were performed to map patterns of phonological errors onto an existing gesture-based model of sound production impairment, with the aim of testing a specific substance-based explanation of phonological processing in aphasic speech production.

The results are incompatible with the assumption that lesions of a phonological neural network interfere with the core knowledge postulated in the Phonological Mind Theory.

### References

- Berent, I. (2013). The phonological mind. *Trends Cogn Sci.*, 17(7), 319-327
- Berent, I., Pan, H., Zhao, X., Epstein, J., Bennett, M. L., Deshpande, V., ... Stern, E. (2014). Language universals engage Broca's area. *PLoS ONE*, 9(4), e95155.
- Tilsen, S. (2016). Selection and coordination: The articulatory basis for the emergence of phonological structure. *Journal of Phonetics*, 55, 53-77.
- Ziegler, W., & Aichert, I. (2015). How much is a word? Predicting ease of articulation planning from apraxic speech error patterns. *Cortex*, 69, 24-39.

## Discrete categories and continuous speech dynamics in typical and atypical speech: How much variability is tolerated by the phonological system?

**Doris Mücke, Anne Hermes, Tabea Thies**

IfL Phonetics, University of Cologne, Germany

[doris.muecke@uni-koeln.de](mailto:doris.muecke@uni-koeln.de)

Chronic deep brain stimulation (DBS) of the nucleus ventralis intermedius (VIM) is an effective treatment for patients with medication-resistant Essential Tremor (ET). However, these patients report that stimulation has a deleterious effect on their speech, severely impacting their quality of life and social functioning (Flora et al. 2010, Mücke et al. 2014). The present study investigates the effect of deep brain stimulation in German Essential Tremor patients on the realisation of prosodic constituents. We recorded 12 ET patients treated with deep brain stimulation in ON and OFF condition and 12 healthy control speakers using 3-D Electromagnetic Articulography. We analysed variability in coordination patterns in syllable onsets with low and high complexity, such as in <Lima> and <Klima> or <Plina>, within the coupling hypothesis of syllable structure (Browman & Goldstein 2000; Nam et al. 2009, Pouplier 2011). Results for the patient group reveal a timing deficit in the phonetic realisation of syllables with high complexity, indicating a lack of complex (non-innate) coupling relations. These timing differences are not categorical but gradient in nature pointing to the fact that there are dynamic mechanisms of regulation behind quantitative consequences of qualitative syllable parses. We discuss how much variability is tolerated in a phonological system before it becomes unstable and patterns of syllable organisation break down. First results point to the fact that most problems arise in postnuclear position: when the level of prominence decreases patients have problems adapting to the requirements of localised reduced speech (compare to Ziegler 2002).

### References

- Browman, C. P., & Goldstein, L. (2000). Competing constraints on intergestural coordination and self-organization of phonological structures. *Bulletin De La Communication Parlée*, 5, 25–34.
- Flora, E.D., C.L. Perera, A.L. Cameron & G.J. Maddern (2010). Deep brain stimulation for essential tremor: A systematic review. *Movement Disorders* 25(11), 1550-1559.
- Mücke, D., J. Becker, M.T. Barbe, T.B. Roettger, I. Meister, L. Liebhart, L. Timmermann, & Grice, M. (2014). The effect of Deep Brain Stimulation on the speech motor system in Essential Tremor Patients. *Journal of Speech, Language, and Hearing Research* 57:1206-1218.
- Pouplier, M. (2011). The atoms of phonological representations. In M. van Oostendorp, K. Rice, B. Hume & C. Ewen (eds.), *The Blackwell Companion to Phonology*. Wiley-Blackwell.
- Nam, H., Goldstein, L., & Saltzman, E. (2009). Self-organization of syllable structure: A coupled oscillator model. In F. Pellegrino, E. Marisco, I. Chitoran, & C. Coupe (Eds.), *Approaches to Phonological Complexity* (pp. 299–328). Mouton de Gruyter, Berlin.
- Ziegler, W. (2002). Task-related factors in oral motor control: speech and oral diadochokinesis in dysarthria and apraxia of speech. *Brain Lang.* 80(3):556-75.

## Spoken language processing across the adult life span

**Esther Janse**

Radboud University, The Netherlands

*Esther.Janse@mpi.nl*

With advancing adult age, language users become more experienced users of their (native) language. This increased language experience, in terms of for instance increased vocabulary knowledge and better semantic networks for older compared to younger adult language users, may make up for possible negative age-related effects of hearing loss, speech motor control and cognitive decline on the production and perception of spoken language. In this talk I will show evidence of stability across the adult life span in the following three language domains: 1) perceptual adaptation for speech comprehension, 2) statistical learning upon presentation of auditory linguistic sequences, and 3) in probabilistic reduction effects in reading aloud. These findings will be discussed in the context of predictive processing, and in the context of situations where age differences are observed.

## When slow speech sounds fast: How the speech rate of one talker influences perception of another talker

**Merel Maslowski<sup>1</sup>, Antje S. Meyer<sup>1</sup>, Hans Rutger Bosker<sup>1,2</sup>**

<sup>1</sup>Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands;

<sup>2</sup>Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands

*Merel.Maslowski@mpi.nl*

Listeners are continuously exposed to a broad range of speech rates. Earlier work has shown that listeners perceive phonetic category boundaries relative to contextual speech rate. This process of rate-dependent speech perception has been suggested to occur across talker changes, with the speech rate of talker A influencing perception of talker B. This study tested whether a 'global' speech rate calculated over multiple talkers and over a longer period of time affected perception of the temporal Dutch vowel contrast /a/-/a:/. First, Experiment 1 demonstrated that listeners more often reported hearing long /a:/ in fast contexts than in 'neutral rate' contexts, replicating earlier findings. Then, in Experiment 2, one participant group was exposed to 'neutral' speech from talker A intermixed with slow speech from talker B. Another group listened to the same 'neutral' speech from talker A, but to fast speech from talker B. Between-group comparison in the 'neutral' condition revealed that Group 1 reported more long /a:/ than Group 2, indicating that A's 'neutral' speech sounded faster when B was slower. Finally, Experiment 3 tested whether talking at slow or fast rates oneself elicits the same 'global' rate effects. However, no evidence was found that self-produced speech modulated perception of talker A. This study corroborates the idea that 'global' rate dependent effects occur across talkers, but are insensitive to one's own speech rate. Results are interpreted in light of the general auditory mechanisms thought to underlie rate normalization, with implications for our understanding of dialogue.

## Unequal effects in the normalization of lexical tone and the normalization of vowel quality

Matthias J. Sjerps<sup>1</sup>, Caicai Zhang<sup>2</sup>, Gang Peng<sup>2</sup>

<sup>1</sup>Donders Institute for Brain Cognition and Behaviour, Donders Centre for Cognitive Neuroimaging, Radboud University Nijmegen, The Netherlands; <sup>2</sup>Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong

[m.j.sjerps@gmail.com](mailto:m.j.sjerps@gmail.com)

Speech from different speakers varies on many dimensions. In tone languages such as Cantonese variation in the acoustic realization of speech sounds can be caused by differences in speakers pitch ranges (affecting lexical-tone height) or by differences in their vocal tract shapes (affecting vowel quality). Listeners perceptually “normalize” (i.e., shift their category boundaries) to a speaker’s expected range of pitch or vowel quality to accommodate such differences (1-4). The current study tested 16 speakers of Cantonese to address the temporal scope over which tone and vowel quality are normalized. Participants heard a nonsense word on every trial (/f?patsi/) where in half of our experiment the first vowel (/ʔ/) ranged between /o/-/u/ (an F1 distinction) in 5 steps while the /patsi/ context was manipulated to have either a high F1 range or a low F1. In the other half of the experiment the first vowel ranged from a low-level tone to a mid-level tone. Here the /patsi/ part was manipulated to have either a high or a low pitch. The results revealed that for normalization of vowel quality only the F1 range of the context on the preceding trials influenced participants’ decisions. For normalization of tone only the pitch range on the immediately following /patsi/ context influenced perception of tone. These results demonstrate that normalization of vowel quality and lexical tone operate over a different temporal scope. These patterns may reflect differences in the underlying mechanisms that cause normalization of tone and of vowel quality.

### References

- 1: Francis, A. L., Ciocca, V., Wong, N. K. Y., Leung, W. H. Y., Chu P. C. Y. (2006). Extrinsic context affects perceptual normalization of lexical tone. *The Journal of the Acoustical Society of America* 119(3) 1712-1726.
- 2: Zhang, C., Peng, G., Wang W. S. Y. (2013). Achieving constancy in spoken word identification: Time course of talker normalization. *Brain and Language* 126(2) 193-202.
- 3: Ladefoged P., Broadbent D. E. (1957). Information conveyed by vowels. *The Journal of the Acoustical Society of America* 29(1) 98-104.
- 4: Sjerps M. J., Reinisch E. (2015). Divide and conquer: How perceptual contrast sensitivity and perceptual learning cooperate in reducing input variation in speech perception. *Journal of Experimental Psychology: Human Perception and Performance* 41(3) 710.

## Long-term effects of accent exposure on perception and production

Helena Levy, Lars Konieczny, Adriana Hanulíková

University of Freiburg, Germany

[helena.levy@frequenz.uni-freiburg.de](mailto:helena.levy@frequenz.uni-freiburg.de)

We examined how school-aged children perceive and produce speech as a consequence of their experience with regional or foreign accents. We know from previous studies that children have more difficulties recognizing words in unfamiliar accents as compared to familiar accents (Nathan et al., 1998; Bent, 2014). Less attention has been paid to the influence of experience with accented speech on accent comprehension. This project is concerned with the effects of type of accent (regional or foreign) and amount of accent experience on children’s comprehension of unfamiliar accents. In a perception experiment, 65 German primary school children (mean age 9 years, 10 months) were asked to repeat sentences spoken by three different speakers: one who spoke standard German, one with a foreign accent (Korean accented German) and one with a regional accent in German (Palatinate German). All of the children had experience with regional and foreign accents but the amount of accent exposure to both kinds of accents varied considerably. More experience with regional accents led to more correct sentence repetitions in the regional accent condition. More experience with foreign accents, however, did not help in the foreign accent condition. Type and amount of experience seem to co-determine processing ease of accented speech. We are currently analyzing the same subjects’ productions of German vowels in spontaneously produced words in order to examine how children’s pronunciation differs acoustically according to type and amount of accent experience. We will discuss the influence of experience with regional and foreign accents on children’s speech perception and production.

### References

- Bent, T. (2014). Children’s perception of foreign-accented words. *Journal of Child Language* 41, 1 – 22.
- Nathan, L., Wells, B., & Donlan, C. (1998). Children’s comprehension of unfamiliar regional accents: a preliminary investigation. *Journal of Child Language* 25, 343 – 365.

## From abstract phonemes to speech movements: the role of orofacial biomechanics and multisensory motor goals.

Pascal Perrier<sup>1</sup>, Jean-François Patri<sup>1,2</sup>, Julien Diard<sup>2</sup>

<sup>1</sup>Univ. Grenoble Alpes, CNRS, GIPSA-Lab, UMR 5216, F-38000 Grenoble, France;

<sup>2</sup>Univ. Grenoble Alpes, CNRS, LPNC UMR 5105, F-38000 Grenoble, France

*Pascal.Perrier@gipsa-lab.grenoble-inp.fr*

In this talk, we will present some results generated with our generic speech production model GEPPETO, that are consistent with the view that investigating physical realizations of phonological units and understanding how they emerge and evolve requires considering crucial physical properties of the speech production apparatus and key-features of speech motor control. First, using 2D and 3D biomechanical models of the tongue and the face, we will show that movement trajectories and then spectral variations that carry phonological information from the speaker to the listener can be significantly influenced by orofacial biomechanics and the variation of its dynamical behavior across speaking conditions (clarity, speaking rate). Second, using our recently developed Bayesian model of speech production planning, we will show how speech variability patterns are constrained by the nature of the motor goals, depending on whether they are auditory, somatosensory or both.

### References

- Perrier P., Payan Y., Zandipour M. & Perkell J. (2003) Influences of tongue biomechanics on speech movements during the production of velar stop consonants: A modeling study. *Journal of the Acoustical Society of America*, 114(3), 1582-1599.
- Buchaillard, S., Perrier, P., & Payan, Y. (2009). A biomechanical model of cardinal vowel production: muscle activations and the impact of gravity on tongue positioning. *Journal of the Acoustical Society of America*, 126(4), 2033-2051.
- Nazari, M.A., Perrier, P., Chabanas, M., & Payan, Y. (2011). Shaping by stiffening: a modeling study for lips. *Motor control*, 15(1), 141-168.
- Patri, J.F., Diard, J., & Perrier, P. (2015). Optimal speech motor control and token-to-token variability: a Bayesian modeling approach. *Biological Cybernetics*, 109 (6), 611–626.

## Nearly-decomposable hierarchies, symbolic grammars and pervasiveness of dynamics

Adamantios Gafos

University of Postdam, Germany

*gafos@uni-potsdam.de*

I present an argument for extending the dynamical approach from representations to principles in grammars. I highlight the key similarities and differences between dynamics in grammars and standard symbolic grammars. I then return to representations by presenting an extension of the dynamical approach to the assembly of phonological form using a task where participants listen to various stimuli as they utter simple syllables; this task zooms in on the assembly of phonological representations at Catford's microchronic time scale. Dynamical principles pervade across the different examples. Why then do grammars seem mostly symbolic? I argue that this fact follows as a consequence of two properties, the omnipresence of hierarchies in nature (and in cognition as emphasized in Smolensky 1988, 2006) and the allied notion of near-decomposability of time scales as promoted in work by Simon (1973) and Rosen (1969).

## Auditory adaptation effects in speech motor control

### Douglas Shiller

École d'orthophonie et d'audiologie, Faculté de médecine, Université de Montréal, Canada

*doug.shiller@gmail.com*

Acoustic speech signals are notoriously variable within and between talkers. To aid in the linguistic decoding of such noisy signals, it is well known that listeners employ a number of perceptual mechanisms – integrating information at multiple levels – to help reduce the impact of linguistically irrelevant acoustic variation. Rapid perceptual accommodation to differences in age and gender is achieved, in part, through vowel-extrinsic normalization, whereby the immediately preceding speech signal provides a frame-of-reference within which talker-specific vowel category boundaries are determined (Ladefoged & Broadbent, 1957). Listeners also draw upon higher-order linguistic information to facilitate phonetic processing of noisy or ambiguous speech acoustic signals, as illustrated by the well-known lexical effect on perceptual category boundaries (Ganong, 1980).

Since their discovery many decades ago, these adaptive perceptual mechanisms have been considered primarily as processes supporting the decoding of ambiguous speech acoustic signals originating from other talkers. Here, I will describe a series of recent studies demonstrating that such perceptual adaptation effects can also alter the processing of self-generated speech acoustic signals (i.e., auditory feedback), and by extension, the sensorimotor control of speech production. The results support the idea that mechanisms of perceptual plasticity, including those driven by higher-level cognitive-linguistic constraints, can interact with the on-line sensory processes guiding speech motor function.

## How do communicative goals affect continuous speech dynamics?

### Caroline Niziolek

Boston University, USA

*carrien@bu.edu*

In this talk, I will present evidence that speakers' communicative goals form the predictions for what their speech should sound like, and that mismatches in prediction activate a neural circuit that helps to catch and dynamically correct impending speech errors. One set of experiments employs altered auditory feedback to drive this detection-correction circuit with an induced "mismatch." Another line of research examines the dynamics of vowel acoustics in the context of natural production variability to show the constant influence of feedback in everyday speech. Through these studies, I will show that the degree of suppression of auditory cortical activity during speech reflects an error detection-correction process that is present even before a full error is realized, and that is modulated by the context-dependent goals of the speaker. My ongoing work with persons with aphasia explores how feedback correction mechanisms can compensate for impaired feedforward speech production, and how we might leverage feedback-based training to develop technologies for speech learning and rehabilitation.

## Adaptation to dichotically presented spectral and temporal real-time perturbations of auditory speech feedback

Mareike Flögel<sup>1</sup>, Christian A. Kell<sup>1,2</sup>

<sup>1</sup>Cognitive Neuroscience Group, Brain Imaging Center, Goethe University, Frankfurt, Germany; <sup>2</sup>Department of Neurology, Goethe University, Frankfurt, Germany

c.kell@em.uni-frankfurt.de

Speech production is influenced by sensory feedback, particularly by the auditory channel. If certain spectral or temporal aspects of the auditory feedback are perturbed in real-time during speech production, the speaker adapts production to compensate for the disturbance<sup>1,2</sup>. Auditory feedback is thought to affect speech production via processes in the right hemisphere.<sup>3</sup> Because, so far, only spectral feedback manipulations were investigated, it is unclear whether the observed right-lateralization reflects a right hemispheric specialization for feedback analyses in general<sup>3</sup> or a right hemispheric specialization for spectral processing<sup>4</sup>. We thus tested whether the adaptation to perturbations of spectral and temporal speech features lateralizes differently. German speakers' auditory feedback was altered spectrally (n=7) or temporally (n=9) during the production of CVC monosyllabic pseudowords. Spectral perturbations increased the vowels' F1 frequency, temporal manipulations decelerated the vowel<sup>5</sup> (20% over 40 trials in steps of 0.05%). Auditory feedback was presented dichotically (feedback manipulation only in one ear while the other ear perceived unperturbed feedback) or diotically (perturbed/unperturbed feedback in both ears). Participants decreased produced vowels' F1 over trials to compensate for the spectral feedback perturbation when perturbed auditory feedback was presented to both ears, or only to the left ear. In contrast, when only the right ear received spectrally altered feedback, no significant change in produced F1 frequency was observed. To compensate for the temporal feedback perturbation, participants shortened their vowel production over trials. For the temporal manipulation, compensation was greatest if altered auditory feedback was presented to both, or only the left ear. These results suggest that spectral features of self-produced speech are monitored more strongly by the right than by the left hemisphere. In contrast, the left more than the right hemisphere processes temporal features of auditory feedback.

### References

- 1 Villacorta, V. M., Perkell, J. S., & Guenther, F. H. (2007). Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. *JASA* 122(4), 2306–2319.
- 2 Mitsuya, T. E., MacDonald, N., & Munhall, K.G. (2014). Temporal Control and Compensation for Perturbed Voicing Feedback. *JASA*, 135(5), 2986–2994.
- 3 Tourville, J. A., & Guenther, F.H. (2011). The DIVA Model: A Neural Theory of Speech Acquisition and Production. *Language and Cognitive Processes* 26(7), 952–81.
- 4 Zatorre, R.J., & Belin, P. (2001). Spectral and Temporal Processing in Human Auditory Cortex. *Cerebral Cortex*, 11(10), 946-953.
- 5 Tourville, J.A., Cai, S. & Guenther, F.H. (2013). Exploring auditory-motor interactions in normal and disordered speech. *Proceedings of Meeting on Acoustics*. 9:060180.

## Individual differences and attentional effects on cue weighting for prosody perception

Jennifer Cole

Northwestern University, USA

jennifer.cole1@northwestern.edu

Through the temporal pattern of F0, intensity, local tempo and other acoustic properties, prosody conveys rich linguistic information. In some languages, prosody encodes discrete lexical contrasts (e.g., tone in Chinese languages), but in many other instances, prosodically encoded meaning is tied to the situational context, marking distinctions that are less clearly categorical (e.g., illocutionary force; referential accessibility). Given the multiplicity of acoustic prosodic cues, and the potential role of contextual factors on listener's perception of prosody, we ask whether listeners agree in their perception of the prosodic features of an utterance, and further, if listeners are similar in the influence that individual cues have on their rating of prosodic features. This investigation is fueled by the broader goal of understanding if and how variability in the perception of prosody impacts the cultural transmission of language, and the role of prosody in linguistic communication. The present study examines individual differences in the perception of prosody in conversational English, as viewed through the lens of prosodic annotation. Prosodic ratings from 32 untrained annotators performing Rapid Prosody Transcription were analyzed with Generalized Additive Mixed Models. Results show that perceived prominences and boundaries are moderately predicted from combined acoustic cues (F0, intensity, duration) and contextual cues (word frequency, POS), but also reveal individual differences among listeners in cue selection and weighting. GAMM findings point to an implicational hierarchy in the selection of cues, where cues with the strongest effect on prosodic rating are selected by the greatest number of listeners, and also a clustering of listeners based on cue selection. Differences in cue weighting are also observed depending on task instructions that focus listeners' attention on the acoustic vs. meaning dimensions of the heard utterance. These findings are discussed in terms of attention and processing strategies used in the perception of prosody.

## Language-independent talker-specificity in bilingual speech production

**Ann Bradlow**

Department of Linguistics, Northwestern University, USA

*abradlow@northwestern.edu*

Current phonetic theories emphasize the integration of linguistic and indexical information. For monolinguals, instance-specific, individual-specific, and group/dialect-specific indexical features are conveyed by phonetic variations of a single language. For bilinguals, language dominance (typically L1 over L2) and structural interactions between the two languages introduce additional sources of variation in the speech encoding of linguistic and indexical information. We probed the relationship between these multiple sources of variation by seeking evidence for talker “trait” characteristics that persist across L1 and L2 in bilingual speech. Using automatically extracted speaking rate measurements (syllables/second) from read and spontaneous speech recordings in the L1 and L2 of bilinguals (n=86, 10 L1s) and in the L1 of English monolinguals (n=27), we examined the relation between talker-specificity and language-specificity as modulated by dominance (L1 vs. L2). We found note-worthy differences in L1 speaking rate across the various languages, and as expected, L2 speaking rates were slower than L1 speaking rates both across groups (monolingual L1 English vs. bilingual L2 English) and within bilinguals across L1 and L2. Critically, L1 speaking rate significantly predicted L2 speaking rate: relatively fast or slow talkers in L1 were also relatively fast or slow in L2, respectively. Similarly, a related study demonstrated that L1 intelligibility predicted L2 intelligibility: relatively high or low intelligibility talkers in L1 were relatively high or low intelligibility talkers in L2, respectively. Together, these results indicate a persistent influence of talker-specific trait characteristics that combine with, rather than are overwhelmed by, language-specific and dominance-dependent influences in bilingual speech.

## Dynamics of lexical access and speech articulation

**Matt Goldrick**

Northwestern University, USA

*matt-goldrick@northwestern.edu*

During speech, the positions of the articulators evolve continuously in physical space. Theories of lexical access have proposed a dynamical process also underlies speech planning. As words and sound structures are retrieved from memory, the state of the mental representations underlying articulation evolve continuously in an abstract symbolic space, organized around the dimensions that define linguistic structure. To support this view, I will review evidence suggesting multiple planning elements can be co-activated and as a result co-produced in articulation. Critically, the degree of co-activation is dynamically modulated by variables affecting lexical access. I'll then discuss recent evidence from my lab extending this work to higher levels of planning; these results suggest that the dynamics of semantically-driven co-activation can modulate articulation.

## Using ultrasound speech imaging to investigate within- and cross-language interference in word production

Martin Corley, Eleanor Drake, Sara Knight

Psychology, PPLS, University of Edinburgh, UK

Martin.Corley@ed.ac.uk

Much psycholinguistic research into language production focuses on naming latencies: That is the time between the presentation of a stimulus and the recording of the onset of acoustic sound. Motor variability is treated as “noise” by-condition differences in response latency are assumed to be expressed only at an abstract cognitive level. However instrumental acoustic analysis of speech output has revealed many speech errors to be non-categorical in nature. Articulatory imaging confirms that far from being noise some of the variability in the speech motor realisations of even perceptually correct responses is systematically related to the presence of representations which compete during the speech production process.

Here we investigate the tongue movements that generate the responses in picture-naming experiments using a novel ultrasound imaging technique to investigate speech motor movement during the naming latency period. We show differences in degree of tongue movement over time across differing participants and experimental conditions. Experiment 1 is based on a classical picture-word interference paradigm and demonstrates that participants make more tongue movements in conditions in which the superimposed word mismatches the named image. Experiment 2 demonstrates that tongue movements in bilingual image naming can be affected by knowledge of another language. In both cases the differences in tongue movements occur early relative to stimulus presentation implicating a language production system in which information cascades quickly from thought to action.

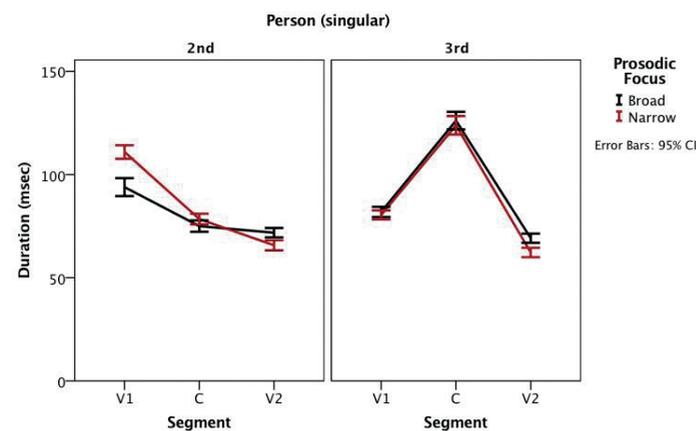
## Temporal information marks person in a Northern Lazio Italian variety

Melissa A. Redford, Sara Pacchiarotti

University of Oregon, USA

redford@uoregon.edu

Castiglione in Teverina (CiT) Italian, like other varieties of Northern Lazio Italian, has *raddoppiamento sintattico* (RS)—gemination at word junctures in running speech. Though originally explained with reference to syntactic constituency (e.g., Nespor & Vogel, 1986), RS is now usually considered purely prosodic. As a prosodic phenomenon, it is subject to blocking. Yet, when raddoppiamento is linked to new meaning it will lexicalize (*a bórdo* “on board” → *abbórdo* “I accost”) and thus become categorical. We believe to have found something like lexicalization of RS in CiT Italian, albeit to signal syntactic relations. In CiT Italian, 2nd and 3rd SG auxiliaries in present perfect constructions are confusable because 2nd person [a] reduces to [a] in running speech where full independent pronouns are also rare. Speakers compensate for this confusability by geminating the onset of the participial verb form in 3rd person; a fact we confirmed experimentally. Our elicitation experiment with 9 CiT speakers further indicated that such gemination is preserved even under narrow prosodic focus, which blocks auxiliary reduction. Specifically, our speakers produced simple clauses with present perfect verb forms first under broad prosodic focus and then in response to a corrective focus prompt. The clauses, repeated 3 times in random order per condition, included 24 different main verbs with initial /f, m, s, n/. Results indicated the expected strong effect of person, but no interaction with focus condition on consonant duration (Figure 1). This result could indicate the stored representation of supralexical units with their own temporal patterning.



**Figure 1.** Initial consonant (C) duration of participial verb form is shown relative to the preceding (auxiliary) vowel (V1) and following unstressed vowel (V2) for the broad and narrow focus conditions. Narrow (corrective) focus was elicited by feigning confusion over whether „he has *verbed*” or “you (fam. sg.) have *verbed*” for the 2nd and 3rd person targets, respectively.

## How your own speech rate can change how you listen to others

**Hans Rutger Bosker**

Max Planck Institute for Psycholinguistics, The Netherlands; Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands

*HansRutger.Bosker@mpi.nl*

In conversation, our own speech and that of others follow each other in rapid succession. However, how the acoustic and indexical properties of self-produced speech interact with speech perception is unknown. This study investigated context effects induced by our own voice through three experiments targeting rate normalization (i.e., perceiving phonetic segments relative to surrounding speech rate).

Experiment 1 revealed that hearing fast and slow context sentences (produced by someone else) alters the perception of subsequent target words, replicating earlier work. Experiment 2 demonstrated that producing fast and slow context sentences *oneself* also alters the perception of subsequent target words (produced by someone else), indicating contextual effects of our own speech rate. Nevertheless, the effect of self-produced speech rate (Experiment 2) was reduced compared to that of perceived speech rate (Experiment 1).

This may be due to indexical properties of our own speech (i.e., talker-incongruency between context and target) or due to the additional task of speech production. Therefore, in Experiment 3, the same participants from Experiment 2 passively listened to their own self-produced context sentences. The effect of contextual speech rate in Experiment 3 was *not* reduced compared to Experiment 1. This suggests that it is *not* the indexical properties of self-produced speech, but rather the additional task of speech production that attenuated the effect of contextual speech rate in Experiment 2, potentially through speaking-induced suppression.

Taken together, this study finds that variation in speech production may induce variation in speech perception, thus carrying implications for our understanding of conversation.

## Exploring the front fricative contrast in Greek: a study of acoustic variability based on cepstral coefficients

Jason Lilley<sup>1</sup>, Angeliki Athanasopoulou<sup>2</sup>, Laura Spinu<sup>3</sup>

<sup>1</sup>Nemours Biomedical Research, USA, <sup>2</sup>Univ. of California San Diego, USA, <sup>3</sup>Univ. of Toronto, Canada

[laura.spinu@utoronto.ca](mailto:laura.spinu@utoronto.ca)

Classification methods based mainly on traditional acoustic measures for fricatives (e.g. spectral peak location, amplitude, and duration) yielded lower success rates (66%) with front fricatives [f, v, θ, ð], compared to 88% with sibilant fricatives [s, z, ʃ ʒ] in English [1]. Perceptual studies show that human listeners also have difficulty distinguishing these sounds, and an acoustic investigation employing both traditional and more innovative measures did not find any cues “even modestly invariant for place of articulation in non-sibilants” [2]. Partially explaining these results, an articulatory study using magnetic resonance imaging showed that labiodental fricatives exhibited the most variability across speakers [3]. In the current study, we explore the factors underlying the difficulty in acoustic classification/perception of front fricatives by taking a closer look at variability in the production of Greek speakers. We apply a novel classification tool based on cepstral coefficients (following [4]) in order to classify front fricatives from an experimental corpus with 25 subjects, and employ statistical methods to classify the tokens from each place of articulation into as many different prototypes as possible. The overall classification rate we obtained is only slightly higher compared to previous work (74%), but we were able to identify distinct subgroups within the four main categories, each with specific characteristics. Our study documents the properties of fricatives in a new language, adds to our understanding of acoustic variability, and sheds more light on the lack of invariance in the production of front fricatives.

### References

- [1] Jongman, A., Wayland, R., and Wong, S. 2000. Acoustic characteristics of English fricatives, *Journal of the Acoustical Society of America* 10/2000; 108 (3 Pt 1), 1252–1263.
- [2] McMurray, B., and Jongman, A. 2011. What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. *Psychological Review* 118, 219–246.
- [3] Narayanan, S., Alwan, A. A., and Haker, K. 1995. An articulatory study of fricative consonants using magnetic resonance imaging. *Journal of the Acoustical Society of America*, 98, 1325–1347.
- [4] Spinu, L. and Lilley, J. 2016. A comparison of cepstral coefficients and spectral moments in the classification of Romanian fricatives. *Journal of Phonetics* 57, 40–58.

## Simulating language change and dynamics – example of Tswana

Jagoda Bruni, Daniel Duran, Grzegorz Dogil

Institut für Maschinelle Sprachverarbeitung Universität Stuttgart, Germany

[jagoda.bruni@ims.uni-stuttgart.de](mailto:jagoda.bruni@ims.uni-stuttgart.de)

In this study we present a hybrid multi-agent modeling framework ‘*Kamoso*’ (from Tswana ‘in the future’) which facilitates investigation into sound change by combining the sociophonetic model of Nettle (1999) and the exemplar-based model of Wedel (2004) into a single unified model. *Kamoso* enables simulation scenarios of different social networks with varying interaction schemes and social distances between the speaking/listening agents. The goal of this framework is to allow examination of competition between different phonetic forms. We illustrate it with the case of phonetically intuitive /mb/ and unintuitive [mp] voicing variants of post-nasal stops in Tswana (unintuitive because it requires more articulatory effort than producing sequences of nasal followed by a voiced stop [Westbury & Keating, 1986]). In our simulation, the structure of an individual agent’s mental lexicon is embedded in an exemplar theoretic setting, whereas the /mb/ → [mp] transition over populations is adapted from Nettle’s Social Impact Theory model (1999) where interacting individuals pass through five life stages before death. The population of agents is embedded within a *social network* which defines social relations between the members of the community including some hyper-influential individuals. Moreover, we assume life-long learning and accommodation to the linguistic environment through continuous adding of the new linguistic experience to the lexicon in each simulation epoch (understood as a life generation). The goal of the simulation is to examine the complex interplay between social influence/bias, frequency of occurrence and functional biases (ease of speech production/discrimination), in how they might account for unintuitive phonetic transition.

### References

- Nettle, D. 1999. Using social impact theory to simulate language change. *Lingua* 108(2-3), 95–117.
- Wedel, A. B. July 2004. Category competition drives contrast maintenance within an exemplar based production/perception loop. *Proceedings of the Seventh Meeting of the ACL Special Interest Group in Computational Phonology Barcelona, Spain. Association for Computational Linguistics*, 1–10.
- Westbury, J. R., Keating, P. A. Mar. 1986. On the naturalness of stop consonant voicing. *Journal of Linguistics* 22(01), 145.

## Producing accented words without self-listening improves word memory

**Ann-Kathrin Grohe, Andrea Weber**

University of Tübingen, Germany

*ann-kathrin.grohe@uni-tuebingen.de*

Own overt word productions facilitate word recall more than listening to others producing the words (MacLeod, 2011). This production advantage applies both for words with an accent that is familiar to participants, and for words with an unfamiliar accent (Grohe & Weber, 2016). It is still possible, however, that hearing oneself while speaking (self-listening) largely drives the effect. The present study tests therefore whether self-produced words are still better recalled when participants do not hear themselves during production. Forty (28 female, 12 male) speakers of the Swabian German accent participated. Test words either had a Swabian accent marker that was familiar to participants (/st/ pronounced as [ʃt], Zahnbür[ʃt]e – ‘tooth brush’) or an unfamiliar Northern German accent (/ft/ pronounced as [st], Blumen[st]rauβ – ‘bouquet’). In the initial training phase, words were presented on a computer screen and participants either read the words aloud with the Swabian or Northern German accent marker, or they listened to the words. While producing words, white noise was played over headphones to prevent participants from hearing themselves. After training, participants decided in a visual memory recognition task for individual words from the training phase (old words) and new words with a button press whether each word was old or new. Hit rates were significantly higher for self-produced words than listened-to words ( $\chi^2=27$ ;  $p<.001$ ); this advantage applied to both the familiar and the unfamiliar accent. Thus, the previously found production advantage is a true production effect: production alone improves recall even when the producers do not hear themselves.

### References

- Grohe, A.-K., & Weber, A. (December 6-9, 2016). Speaking aloud improves recall for words with familiar and unfamiliar accents. *Poster presented at the 16th Australasian Conference on Speech Science and Technology*, Sydney, Australia.
- MacLeod, C. M. (2011). I said, you said: the production effect gets personal. *Psychonomic Bulletin & Review*, 18(6), 1197–1202.

## Minimum time for producing and perceiving pitch directions by tonal and non-tonal speakers

**Si Chen<sup>1</sup>, Yiqing Zhu<sup>2</sup>, Ratee Wayland<sup>2</sup>**

<sup>1</sup>Department of Chinese and Bilingualism, The Hong Kong Polytechnic University, Hong Kong; <sup>2</sup>Department of Linguistics, University of Florida, USA

*sarachen@polyu.edu.hk*

Previous studies examined time needed to produce a rising or falling pitch direction. They demonstrated that non-tonal speakers are faster in producing pitch directions than tonal speakers, despite the latter group's linguistic experience with local pitch changes. The study is the first to compare production and perception of pitch directions between the two groups. We created continua of fundamental frequency on different vowels and with different duration, based on estimation of parameters from a Mandarin corpus. Fifteen native Mandarin speakers and fifteen native English speakers were recruited from the Hong Kong Polytechnic University and the University of Florida. Within each duration, we recorded the estimated semitone where the identification rate is 0.5, assuming that the cut-off semitone values are the smallest which a rising or falling pitch direction can be perceived as different from a level tone. Formulas were proposed for perception time needed for effectively pitch directions on low and high vowels by tonal and non-tonal listeners. Non-tonal listeners need longer time to perceive a rising pitch direction less than 10 semitones and a falling pitch direction less than 12 semitones than tonal listeners, but they need less time otherwise. Vowel quality also affects duration needed for effectively perceiving rising and falling pitch directions.

## The effect of one's own voice and production skills on word recognition in a second language

**Nikola Anna Eger, Eva Reinisch**

Institute of Phonetics and Speech Processing, Ludwig Maximilians University Munich, Germany

*eger@phonetik.uni-muenchen.de*

Second language (L2) learners often do not overcome their foreign accent even after long years of training and despite their ability to notice the accent of fellow learners. We asked whether this could be because learners are less aware of their own accent, possibly due to high familiarity with their own acoustic and articulatory speech patterns. The prediction that follows is that learners are better at understanding accented words if they were produced themselves than by other, equally proficient learners. This was tested in two perception experiments.

24 German learners of English were recorded producing a randomized list of difficult minimal word pairs. Based on the produced acoustic differences between the words of each pair, learners were assigned to one of three proficiency groups. Several weeks later, participants returned for a perception task in which they had to listen to the words in isolation and reconstruct the intended word of the minimal pair. They were presented their own as well as others' productions matched in proficiency. The best and worst speakers were also asked to reconstruct the words of the "opposite" proficiency group.

Results showed a self benefit for all proficiency groups relative to other speakers matched in proficiency. Moreover, all participants showed a perceptual benefit if high proficient productions were heard, but this effect was stronger for those who were themselves highly proficient. We conclude that L2 production and perception abilities are linked and that adaptation to one's own accent could indeed lead to lower awareness in recognizing one's own errors.

## The not-so-selective adaptation of vowels

**Miquel Llompart, Eva Reinisch**

Institute of Phonetics and Speech Processing, Ludwig Maximilians University Munich, Germany

*M.Llompart@phonetik.uni-muenchen.de*

Previous research points towards differences between the perception of consonants and vowels. The present study contributes to characterizing these differences by examining the low-level perception of vowels using selective adaptation. Selective adaptation is the effect that after repeated exposure to one endpoint of a sound continuum, more sounds along the continuum will be categorized as the opposite category. Previous studies using this paradigm focused on consonants and demonstrated limited generalization across acoustic and articulatory dimensions such as place of articulation. For example, the categorization of a [ba]-[da] continuum is not affected by repeated exposure to [ga] which is articulated further back than the „back“ continuum endpoint (i.e., [da]). Motivated by the fact that vowels are produced and perceived in a more gradient manner than consonants, we tested whether generalization could be found with vowels that were presented in adaptor words. Specifically, we tested selective adaptation of a vowel height continuum ([u]-[o]) and a place of articulation (front-back) continuum ([y]-[u]) with the endpoint adaptors and other adaptors that only matched in one dimension. For example, [e] has the same height as [o], and [i] is front like [y]. Results showed selective adaptation of the same magnitude for all adaptors in the direction of the endpoint that matched in the critical dimension. Although the cause of this effect remains unclear, as it could be either adaptation of phonological features or simple acoustic contrast, our findings provide evidence supporting the recurrent claim that vowel perception is less restrictive than the perception of consonants.

## On the perception of voicing in whisper: a cross-modal semantic priming study

Yohann Meynadier, Sophie Dufour

Aix Marseille Univ, CNRS, LPL, Aix-en-Provence, France

yohann.meynadier@lpl-aix.fr

This study focuses on the perception of the voicing feature in whispered words in French. A voiced consonant in whispered voice is produced without vibration of the vocal folds, i.e. the main property of the [+voiced] feature in this language. In French, some studies show that the [+voice] whispered obstruents retain some phonetic properties of their underlying identity, regarding acoustical traces related to their duration (Vercherand, 2010, Meynadier et Gaydina, 2013), to their intra-oral pressure (Meynadier, Gaydina 2013, Garnier et al. 2014, Meynadier 2015) or to their glottal area (Malécot et Peebles 1965, Crevier-Buchman et al. 2009, Meynadier 2015). However, even fewer studies teach us how the voicing contrast could be recognized by French. The few studies examining how French listeners perceive the voicing contrast in whispered speech have produced elliptical and contradictory results with not really comparable methods (Vercherand 2010, Fux 2012, Meynadier et al. 2013). Here, the perception of the voicing feature of whispered obstruent consonants is examined in two cross-modal semantic priming experiments. In Experiment 1 with visual targets presented at the offset of auditory primes, a priming effect of similar magnitude to that observed in modal voice is found only when the whispered prime includes a voiceless obstruent (e.g. [s] in *dessert* primes CHOCOLAT). No priming effect appears when the whispered prime includes a voiced obstruent (e.g. [z] in *désert*), neither on the target word SABLE (sand) semantically related to *désert*, nor on the target word CHOCOLAT semantically related to *dessert*. In Experiment 2 with visual targets presented 50 ms after the offset of auditory primes, primes with a whispered voiced obstruent facilitate the processing of their respective semantic associated target words (i.e. [z] in *desert* primes SABLE).

Hence, our study shows that the reconstruction of the voiced feature is not immediate during whispered word recognition and requires a certain amount of time. During this time, the listener may extract the phonetic traces needed to recover the underlying voicing of whispered voiced obstruents.



### References

- Crevier-Buchman L., Vaissière J., Henrich N., Vincent C., Hans S., Brasnu D. (2009). Laryngeal behavior in whispered voice: a study using high speed imaging. *The Voice Foundation's 38th Annual Symposium: Care of the Professional Voice*. Philadelphia.
- Fux T. (2012). *Vers un système indiquant la distance d'un locuteur par transformation de sa voix*. Unpublished PhDissertation, Université de Grenoble.
- Garnier M., Bouhake S., Jeannin C. (2014). Efforts and coordination in the production of bilabial consonants. *Proceedings of the 10th Int. Seminar on Speech Production*, 138-141. Cologne.
- Meynadier Y. (2015). Aerodynamic tool for phonology of voicing. *18th ICPHS*, paper#0497. Glasgow.
- Meynadier Y., Dufour S., Gaydina, Y. (2013). Duration as perceptual voicing cue in whisper. *The 6th Phonetics and Phonology in Iberia Conference*. Lisbon.
- Meynadier Y., Gaydina, Y. (2013). Aerodynamic and durational cues of phonological voicing in whisper. *Proceedings of the 14th Interspeech*, 335-339. Lyon.
- Malécot A., Peebles K. (1965). An optical device for recording glottal adduction-abduction during normal speech. *Zeitschrift Für Phonetik, Sprachw. Und Kommunikations*. 18, 545-550.
- Vercherand G. (2010). *Production et perception de la parole chuchotée en français: analyse segmentale et prosodique*. Unpublished PhDissertation, Université Paris VII.

## How much abstraction is too much? The case of gemination

**Holger Mitterer**

Department of Cognitive Science University of Malta, Malta

*holger.mitterer@um.edu.mt*

Many languages distinguish short and long consonants, or singletons and geminates. Most current proposals on how to phonologically represent geminates share one assumption: The difference between singletons and geminates is relatively uniform for all consonants in a given language. Perception of geminates would then be based on the recognition of a given consonant plus the recognition of a gemination feature or timing unit. It was tested whether this is feasible in Maltese. A production study showed that cues for gemination differ strongly between consonants. For oral stops and fricatives, duration is the strongest cue, with little variation in how strong a segment is produced (measured as voicing leak in voiceless segments). However, for the glottal stop and /h/, different allophones are used for singletons and geminates. A singleton glottal stop mostly surfaces as glottalization while the geminate is mostly produced as a stop. For /h/, the singleton is produced as breathiness on the vowel while the geminate is produced as a back fricative (/x/ or /χ/). A perception study showed that these secondary cues for glottal consonants are quite strong and influence perception not only when consonant duration is ambiguous but also – and in contrast to earlier studies – when the consonant has a typical duration for a singleton or geminate, while, for oral consonants, duration has more leverage. This indicates that geminates may take on irreducible properties and may become segments in their own right, to some extent independent of the respective singletons.

## C-to-V coarticulation in spontaneous French: acoustic analysis and classification

**Fanny Guitard-Ivent, Cécile Fougeron**

Laboratoire de Phonétique et Phonologie (CNRS/UMR7018), Université Sorbonne-Nouvelle Paris 3, France

*fanny.ivent@univ-paris3.fr*

Various sound changes originate from synchronic variations (Ohala, 1981, 1993; Recasens, 2014). Harrington (2012) showed that in German high back vowels are more likely to front synchronically in coronal contexts than front vowels are likely to retract, giving a possible explanation to the origin of front rounded vowels. Here, we test this tendency in French by looking at the coarticulatory effect of front and back consonants on the spectral properties of French oral vowels. More than 21.000 vowels occurring in a C1VC2 context are extracted from a large corpus of casual French recorded in natural conditions (Torreira et al. 2010). V is one among the 9 following oral vowels (/i,y,e,ɛ,ə,a,u,o,ɔ/ with /ə/=/ø, œ, ə/) and C1 & C2 vary between labial, coronal, velar or uvular place of articulation (/labV{cor, vel, uvul}/ & /{cor, vel, uvul}Vlab/). A strong fronting effect in coronal contexts is observed for back vowels (high and non-high). Interestingly for the front vowels, while high front vowels show little retraction, a strong coarticulatory effect is found for the non-high front vowels in /R/ context. This comprehensive examination of contextual influence in this large dataset highlights the large amount of overlap between vowel categories in the acoustic space. For instance, between /a/ and /ɛ/ or between /ə/ and /ɔ/, the last pair being described as a possible merger in French. Results from an ongoing automatic acoustic classification, testing the discriminability of vowels according to the consonantal context, will be presented at the conference.

### References

- Ohala, J. J. 1981. The listener as a source of sound change, in Papers from the Parasession on Language and Behavior, edited by C. S. Masek, R. A. Hendrick, and M. F. Miller *Chicago Linguistic Society*, Chicago, pp. 178–203.
- Ohala, J. J. 1993. The phonetics of sound change. in *Historical Linguistics: Problems and Perspectives*, edited by C. Jones Longman, London, pp. 237–278.
- Harrington, J. 2012. The coarticulatory basis of diachronic high back vowel fronting, in M., Solé, J., Recasens, D., *The Initiation of Sound Change: Perception, Production and Social Factors*, pp. 103-122.
- Recasens, D. 2014. *Coarticulation and Sound Change in Romance*. Amsterdam: John Benjamins Publishing Company.
- Torreira, F., Adda-Decker, M., Ernestus, M. 2010. The Nijmegen corpus of casual French, *Speech Communication*, 52, pp. 201–212.

## Rhythmic convergence during conversational interactions: preliminary evidence in Spanish-speaking dyads

Leonardo Barón Birchenall, Noël Nguyen

Aix Marseille Univ, CNRS, LPL, Aix-en-Provence, France

leonardo.baron@lpl-aix.fr

In some theoretical frameworks, an increase of similarity in speech patterns, known as phonetic convergence, occurs automatically between speakers during conversations (e.g. Louwerse et al., 2012). The present work focuses on the interaction between speech rhythm and phonetic convergence in an interactive task. Specifically, given that repetitions of a speech stimulus can reduce the time and neural activation needed for its processing, and that multiple repetitions significantly enhance memory and learning (Falk et al., 2014), we propose that the use of regular rhythmic structures during conversations produces more convergence between speakers with respect to irregular rhythmic structures. We created a set of stimuli consisting of seven groups of 16 nine- or eight-syllable Spanish sentences each. Each group had a particular rhythmic structure, obtained through the arrangement of different types of words (oxytones, paroxytones, proparoxytones and unstressed words) in feet of different length. Rhythmic structures were composed as follows (unstressed syllables are represented by a lowercase x and stressed syllables by an uppercase X): regular structures (xxXxxXxxX, XxxXxxXxx, XxXxXxXx), irregular structures (xxXxXxxXx, XxXXxxxXx, XxXXxxXx, XxxxXxxXx). We tested four dyads of Spanish native speakers separately in a reading - repetition task, with different combinations of the rhythmic structures. A rhythmic distance score, proposed by Späth et al. (2016), served to determine the degree of convergence between the interlocutors' rhythms. Results indicate a greater amount of convergence between regular structures with respect to irregular ones, when feet nuclei are left aligned. Detail will be given on the response patterns observed in the other conditions, and implications for current models of phonetic convergence in speech will be discussed.

### References

- Falk, S., Rathcke, T., & Dalla Bella, S. (2014). When speech sounds like music. *Journal of Experimental Psychology: Human Perception and Performance*, 40(4), 1491-1506.
- Louwerse, M., Dale, R., Bard, E., & Jeuniaux, P. (2012). Behavior matching in multimodal communication is synchronized. *Cognitive Science*, 36(8), 1404-1426.
- Späth, M., Aichert, I., Ceballos, A., Wagner, E., Miller, N., & Ziegler, W. (2016). Entraining with another person's speech rhythm: Evidence from healthy speakers and individuals with Parkinson's disease. *Clinical Linguistics and Phonetics*, 30(1), 68-85.

## Reciprocal metathesis: the transposition of two non-adjacent segments

Ander Egurtzegi

Institute of Phonetics and Speech Processing, Ludwig Maximilians University Munich, Germany

egurtzegi@phonetik.uni-muenchen.de

In this paper, I present an outline of the little-studied reciprocal metathesis (Utan 1978), which refers to the cases of metathesis that involve two different non-adjacent segments that exchange their positions with one another without directly affecting the rest of the sequence. This kind of metathesis is exemplified by examples such as Lat. *leriquiae* < Lat. *rēlīquīae* 'relic', Fr. *moustique* 'mosquito' (cf. Lat. *musca* 'fly') and non-standard English *irrevelant* < standard English *irrelevant*. Although this process is not as widely discussed in the literature as other kinds of metathesis, it has been known since the early 20th century (cf. Brugmann 1904 [1970]).

In order for reciprocal metathesis to occur, the two affected segments typically need to be in the same syllabic position – i.e., onset, nucleus or coda – and they need to share some crucial phonological features. In languages such as Basque or Greek, reciprocal metathesis affects not only consonants, as in Basque *bage* > *gabe* 'without' or *erakutsi* > *eratsuki* 'show', but vowels as well, as shown by *atera* > *etara* 'come out' or *alkandora* > *alkondara* 'shirt' (Egurtzegi 2014). This paper presents examples of reciprocal metathesis from a wide variety of languages, including Greek, Spanish, Sardinian, Polish, Saraiki and Basque. I hypothesize that, unlike other kinds of metathesis such as perceptual metathesis (Blevins & Garrett 2004), reciprocal metathesis originates in motor planning errors, being similar to the speech error usually regarded as spoonerism (MacKay 1970), in which the sequential order of two segments is involuntarily reversed (cf. *overinflated state* → *overinstated flate*; *pup pocket* → *pos pucket*; Goldstein 1968). These speech errors are the consequence of the influence of some planning elements —such as gestures— in others through priming, coactivation, inhibition, etc. (Garrett & Johnson 2013). I propose that reciprocal metathesis involves cases of gesture exchange errors that can be incorporated into a language, thus yielding sound change. This hypothesis implies that not only segments, but also lower units in the speech chain can exchange their sequential positions with one another as an instance of reciprocal metathesis at the feature level. This is the case in Standard Polish *izdepka* > North Mazovian Polish *izbetka* 'room (dim.)' and Hindi *biṭṭī*: to Saraiki (Western Punjabi, Indo-Aryan) *piḍḍi*: 'small'.

This study aims to build on the typology of metathesis presented by Blevins & Garrett (2004) by adding a different kind of phonetically driven metathesis, which can be understood under the same assumptions accepted for any other phonetic process.

#### References

- Blevins, J., & Garrett, A. 2004. The evolution of metathesis. In Hayes, B., Kirchner, R. & Steriade, D. (Eds.), *Phonetically based phonology*. Cambridge: CUP, 117–156.
- Brugmann, K. 1904 [1970]. *Kurze vergleichende Grammatik der indogermanischen Sprachen*. Strassburg: Trübner [Berlin: Walter de Gruyter].
- Egurtzegi, A. 2014. *Towards a phonetically grounded diachronic phonology of Basque*. PhD Dissertation, University of the Basque Country.
- Garrett, A., & Johnson, K. 2013. Phonetic bias in sound change. In Yu, A.C.L. (Ed.), *Origins of sound change: Approaches to phonologization*. Oxford: Oxford University Press, 51–97.
- Goldstein, M. 1968. Some slips of the tongue. *Psychological Reports* 22, 1009–1013.
- MacKay, D.G. 1970. Spoonerisms: The structure of errors in serial order of speech. *Neuropsychologia* 8, 323–350.
- Ultan, R. 1978. A typological view of metathesis. In Greenberg, J. (Ed.), *Universals of human language* (vol. 2). Stanford, CA: Stanford University Press, 368–399.

## What native language can and cannot do: perception of onset consonant clusters

Harim Kwon<sup>1</sup>, Ioana Chitoran<sup>1</sup>, Marianne Pouplier<sup>2</sup>, Tomas Lentz<sup>2</sup>, Takeki Kamiyama<sup>3</sup>, Philip Hoole<sup>2</sup>

<sup>1</sup>Université Paris Diderot, France; <sup>2</sup>Ludwig-Maximilians-Universität Munich, Germany;

<sup>3</sup>Université Paris 8, France

*harim.kwon@univ-paris-diderot.fr*

Cross-language research on consonant cluster production has shown that consonant clusters in different languages are produced with different degrees of articulatory timing lag. For instance, German consonant clusters are produced with relatively shorter lag between two consonants than Georgian ones [1, 2]. This study examines perceptual sensitivity to these cross-linguistic timing differences in consonant clusters. Native listeners of Georgian, German, French, and Japanese are tested on an AXB similarity judgment test using stimuli including consonant clusters produced by German and Georgian speakers. Stimuli are /bla, gla, gna/ syllables recorded along with articulatory (EMA) data. Short lag German tokens and long lag Georgian tokens are selected as A and B, with Xs of varying degrees of lag chosen from either Georgian or German recordings. Results show that all four groups of listeners are sensitive to the cross-linguistic differences in articulatory timing lag: when the timing lag of X is closer to A (or B), participants are more likely to choose A (or B, respectively). [Georgian:  $\beta = 0.69$ ,  $p < 0.001$ ; German:  $\beta = 0.74$ ,  $p < 0.001$ ; French:  $\beta = 0.44$ ,  $p < 0.001$ , Japanese:  $\beta = 0.48$ ,  $p < 0.01$ ]. This finding suggests that adult listeners are capable of discerning non-native subphonemic details regardless of their native phonotactics. Effects of sub-phonemic details on similarity judgments by the four listener groups are investigated. These include different measures of articulatory lags and acoustic properties that are known to be related to inter-gestural timing within clusters (e.g., duration of vocalic release). Implications of possible task effects will be discussed.

#### References

- [1] Bombien, L., Mooshammer, C. & Hoole, P. 2013. Articulatory coordination in word-initial clusters of German. *Journal of Phonetics*, 41, 546-561.
- [2] Chitoran, I., Goldstein, L. & Byrd, D. 2002. Gestural overlap and recoverability: Articulatory evidence from Georgian. In C. Gussenhoven, T. Rietveld, and N. Warner (eds.) *Laboratory Phonology* 7. Berlin: Mouton de Gruyter. 419-447.

## The development of vowel space in German speaking children

Stella Krüger<sup>1</sup>, Jan Ries<sup>1</sup>, Aude Noiray<sup>1,2</sup>

<sup>1</sup>Laboratory for Oral Language Acquisition (LOLA), University of Potsdam, Germany;

<sup>2</sup>Haskins Laboratories, New Haven, USA

stellkrue@uni-potsdam.de

Vowel acquisition involves establishing static articulatory configurations (paradigmatic aspect) and their integration into coarticulated speech (syntagmatic aspect). There is broad consensus that children by the age of three are capable of producing adult-like isolated vowels suggesting that the paradigmatic aspect is mastered by this age (James et al., 2001).

Although vowel productions may be perceptually intelligible there is still refinement on the level of syntagmatic aspects due to maturational processes of speech motor control and anatomical growth leading to variability in vowel production (Ménard et al., 2007; Vorperian & Kent, 2007). This study examines the maturation of the vowel space in a cross-sectional survey of German children aged 3y, 4y, 5y, 7y and adults.

Rounded and unrounded tense long vowels (/i:/, /y:/, /u:/, /a:/, /e:/, /o:/) were collected in non-words of the structure C1VC2ə. We recorded both the acoustic speech signal and movement of the tongue with ultrasound imaging. Our sample included data from boys and girls to account for both age related changes and sexual dimorphism. The main first three formants were extracted at the midpoint of the vowel semi-automatically using Praat script (Ménard et al. 2007) and compared with manual inspection.

The 3-year olds are expected to show the greatest within and between-speaker variability for all three formants. With increasing age and development of speech motor control, both variability and formant frequency should decrease. We expect to observe effects from sexual dimorphism by the age of 4 years. The data are currently analyzed.

### References

- James, D., van Dorn, J. & McLeod, S. (2001). Vowel production in mono-, di- and poly-syllabic words in children 3;0 – 7;11 years. In L. Wilson & S. Hewat (Eds.), *Proceedings of the Speech Pathology Australia Conference*, 127–136.
- Ménard, L., Schwartz, J. L., Boë, L. J., & Aubin, J. (2007). Articulatory–acoustic relationships during vocal tract growth for French vowels: Analysis of real data and simulations with an articulatory model. *Journal of Phonetics*, 35(1), 1-19.
- Vorperian, H. K., & Kent, R. D. (2007). Vowel acoustic space development in children: A synthesis of acoustic and anatomic data. *Journal of Speech, Language, and Hearing Research*, 50(6), 1510-1545.

## Exposure-induced and training-induced learning of distorted speech: Does learning fail those who need it most?

Karen Banai, Limor Lavie

Department of Communication Sciences and Disorders, University of Haifa, Israel

kbanai@research.haifa.ac.il

Speech perception improves substantially with practice, but the role of perceptual learning (PL) in speech processing remains debated. One hypothesis is that PL serves to rapidly normalize the perception of non-optimal (e.g., accented, rapid) speech. This study was designed to test a facet of this hypothesis, that rapid perceptual learning of degraded speech is weaker in populations with difficulties in speech perception, in this case older adults and older adults with age-related hearing loss. Rapid PL was observed in older adults with normal hearing (NH, n=52) and in those with hearing loss (HL, n=36). However, the magnitude of learning in the two groups of older adults was smaller than in young adults (YA, n = 55) even when starting performance was equated across groups (Cohen's d = 1.5, 0.7 and 0.4 in the YA, NH and HL groups respectively). Additional training yielded learning in all three groups. Nevertheless, like rapid learning, practice-induced learning and its generalization to untrained time-compressed tokens were weaker in the two groups of older adults than in the young adults. We suggest that aging contributes to declines in perceptual adjustments to distorted speech with an additional smaller effect for hearing loss. Further training is not sufficient to offset this decline.

## Introducing the Ventriloquist paradigm: combining dialogue with full control over phonetic detail

**Mirjam Broersma**

Centre for Language Studies, Radboud University, Nijmegen, The Netherlands

*m.broersma@let.ru.nl*

The novel Ventriloquist paradigm enables the study of phonetic accommodation and perceptual learning in dialogue, while allowing full control over the phonetic detail of the input that participants are exposed to. It is being developed to investigate L2 speech learning for production and perception in an ecologically valid yet maximally controlled way. Participants take part in a dialogue which they believe to be genuine; in fact, however, their (real-life) interlocutor is a confederate whose speech is not just 'scripted' (as in the confederate scripting task), but fully prerecorded. This guarantees control over all characteristics of the speech input, e.g. precluding that a confederate accommodates his/her pronunciation to the participant. The set-up is fully tuned to upholding the illusion that the confederate is actually speaking with the participant. The confederate sits opposite the participant, face briefly hidden when he/she "speaks". Participants hear the prerecorded speech over closed headphones. In addition to the standard input, to facilitate a smooth flow of the conversation, the confederate can play prerecorded non-verbal interaction markers and stop-gap replies to any unanticipated remarks or questions from the participant. The new paradigm thus reconciles ecological validity with experimental control for the study of (L2) phonetic processing in dialogue.

## Asymmetric representations of lexical pitch

**Yuki Asano**

University of Tübingen, Germany

*yuki.asano@uni-tuebingen.de*

This study explores underlying representations of lexical pitch in Japanese and Chinese. Although both languages exhibit lexical pitch, Japanese is classified as a pitch-accent language and Chinese a tone language (Hyman, 1978). Tonal references in Japanese are claimed to be privative, i.e. are solely /H(igh)/ (and /Ø/), resulting in asymmetric representations, while those in Chinese are /H/ and /L(ow)/ (Hyman, 2009), i.e. symmetric representations. A speeded lexical decision task using minimal pair words and pseudo-words that contrasted in pitch (HL with a pitch accent on the first mora vs. LH without pitch accent in Japanese/Exp.1, and HL with the 2nd tone and LH with the 4th tone in Chinese/Exp.2) was completed by 24 native listeners of each language. Reaction times and response accuracy were analyzed using linear mixed effects models.

The results of Japanese/Exp.1 showed a significant interaction between *pitch* (HL vs. LH) and *wordness* (word vs. non-word), both in reaction times and response accuracy (both  $p < 0.001$ ); LH words were more difficult to process than HL words, but this did not apply to non-words, see Figure 1.

The finding provides empirical evidence that the tonal reference represented in the Japanese mental lexicon is /H/ and asymmetric without /L/. Exp.2 is currently being conducted, and the results comparing Japanese and Chinese will be presented at the conference. The study expands current models of spoken word recognition, incorporating still understudied lexical prosody and its representations.

### References

- Hyman, L. M. (1978). Tone and/or accent. In D.J. Napoli (ed.), *Elements of tone, stress, and intonation*. Washington, D.C.: Georgetown University Press.
- Hyman, L. M. (2009). How (not) to do phonological typology: the case of pitch-accent. *Language Science* 31, 213-238.

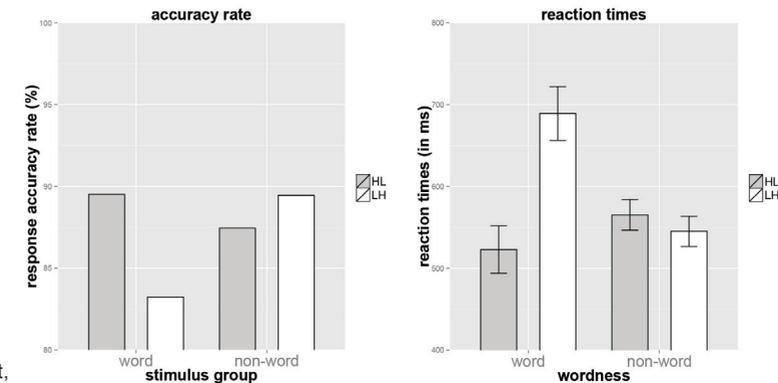


Figure 1. Mean accuracy rates and RTs with 95% CI bars (Exp.1).

## Tracking the timecourse of perception: Ambiguous rhoticity in Glasgow

Robert Lennon, Jane Stuart-Smith, Rachel Smith

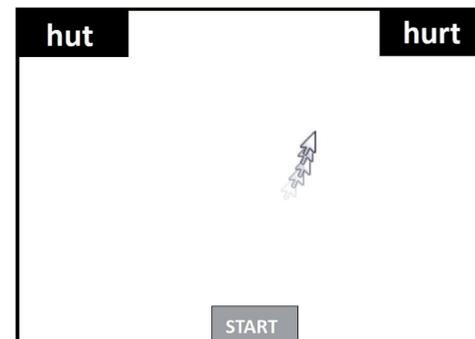
Glasgow University Laboratory of Phonetics, UK

r.lennon.1@research.gla.ac.uk

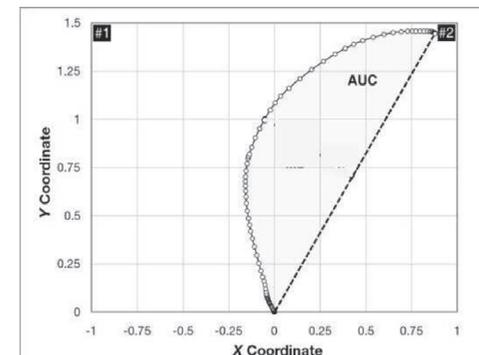
Working class (WC) Glaswegian speech is losing postvocalic /r/ [1], while middle class (MC) rhoticity strengthens [2]. WC minimal pairs like *hut/hurt* can be misperceived, as the pharyngealized /r/ is acoustically [2] and perceptually [3] similar to the preceding vowel. Misperception decreases as long-term experience of Glaswegian increases, and short-term exposure promotes perceptual change [3]. This paper uses mouse tracking [4] to investigate the timecourse of perception as the word unfolds, recording cursor trajectories as participants move the mouse to select e.g. 'hut' or 'hurt' (Fig.1). Trajectory analyses quantify spatial attraction to competitors (Area-Under-the-Curve: AUC, Fig.2 [4]), and curve properties (Discrete Cosine Transforms: DCT [5]). 51 Glaswegians heard 192 stimuli (1xMC and 1xWC Glasgow males) in 3 blocks (MCblock&WCblock alternated their presentation order by participant):

1. MCblock: 24xMC (e.g. *hut/hurt*) + 24xMC distractors (e.g. *meeek/make*)
2. WCblock: 24xWC + 24xWC distractors
3. Rblock: All stimuli from MC&WCblocks, Randomised together.

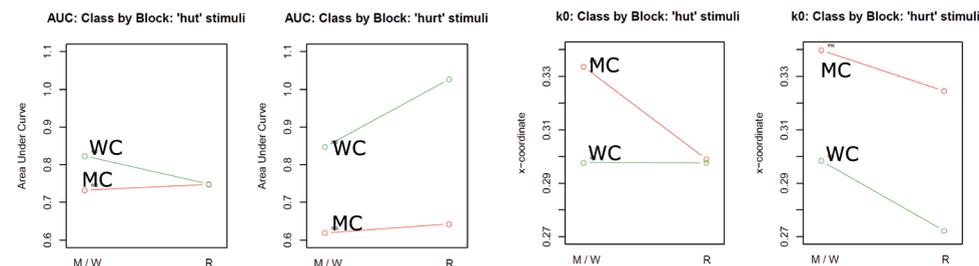
Using lme4 & lmerTest's step() to refine saturated mixed-effects-models in R, the best-fitting AUC model contained the 3-way-interaction of Class:Block:Coda (Coda=whether stimulus canonically has an /r/) ( $Pr(>F)=0.01$ ,  $F=6.02$ ) (Fig.3). Trajectories for WC *hurt* stimuli in the Rblock had the largest AUC (i.e. largest spatial attraction to competitor), suggesting that they were hardest to distinguish from hut. The best DCT model for k0 (mean-x-coordinate/time) also contained Class:Block:Coda ( $Pr(>F)=0.004$ ,  $F=8.51$ ) (Fig.4). With the lowest mean-x-coordinate/time (i.e. latest movement towards correct response), WC *hurt* stimuli were again hardest. Furthermore, all MC stimuli were harder to distinguish when randomised with WC stimuli (Rblock), suggesting a similar finding as [6], highlighting the difficulty of perceptually switching between different speakers with different accents.



1: Mouse Tracker[4]



2: AUC[4]



3: AUC Class:Block:Coda interaction  
(WC=green, MC=red)

4: DCT, k0(mean x-coordinate/time)  
Class:Block:Coda interaction (WC=green,  
MC=red)

### References

- [1] Stuart-Smith, J. (2007). A sociophonetic investigation of postvocalic /r/ in Glaswegian adolescents. *ICPhS 16*, Saarbrücken, Germany.
- [2] Lennon, R., Smith, R., & Stuart-Smith, J. (2015). An acoustic investigation of postvocalic /r/ variants in two sociolects of Glaswegian. *ICPhS 18*, Glasgow, Scotland.
- [3] Lennon, R., Smith, R. & Stuart-Smith, J. (2016). Ambiguous rhoticity in Glasgow: Short term exposure promotes perceptual adaptation for experienced and inexperienced listeners. *LabPhon15*, Ithaca, USA.
- [4] Freeman, J. B., & Ambady, N. (2010). MouseTracker: Software for studying real-time mental processing using a mouse-tracking method. *Behavior Research Methods*, 42(1), 226-241.
- [5] Watson, C. I., & Harrington, J. (1999). Acoustic evidence for dynamic formant trajectories in Australian English vowels. *Journal of the Acoustical Society of America*, 106(1), 458-468.
- [6] Mullennix, J. W., & Pisoni, D. B. (1990). Stimulus variability and processing dependencies in speech perception. *Perception & Psychophysics*, 47(4), 379-390.



## Probing the representation of temporal structure in speech using auditory feedback manipulation

Philip Hoole, Laura Sichlinger

Institute of Phonetics and Speech Processing, Munich University, Germany

hoole@phonetik.uni-muenchen.de

Compared to real-time manipulation of formants, much less is known about whether subjects' sensitivity to discrepancies between predicted and actual feedback extends to the temporal structure of speech. Here we follow an approach pioneered by Cai et al. (2011; see also Tourville et al., 2013). Unlike their study, which used very vocalic material, we explore here the possibility of applying temporal perturbations to different locations in the syllable in items with complex syllable structure. We refer to the key experimental conditions as the onset and coda conditions. In the onset condition, using the German word "Schnecken", the onset consonants /ʃn/ of the first syllable were stretched by 50%, and the following vowel correspondingly compressed, so that the overall duration of the word remained unchanged. In the coda condition, using the word "Menschen", the vowel in the first syllable was stretched by 50%, and the coda consonants /nj/ compressed. Based, among others, on considerations in the coupled-oscillator model of syllable structure (Goldstein & Pouplier, 2014) we hypothesized that the onset manipulation would lead to clearer responses.

As a further new feature we used a paradigm in which the perturbation was designed to become predictable for the subjects over the course of the experiment, i.e. following a baseline phase the perturbation was gradually increased over a ramp phase (this in turn followed by a hold phase with maximum perturbation, and an after-effect phase with normal feedback). This was designed to determine whether subjects are in principle able to compensate for a temporal perturbation by adjusting the durations of the perturbed segments themselves. Thus analysis focused on segmental durations in the hold phase vs. the baseline phase. In the onset condition (initial cluster stretched, following vowel compressed) there was no difference in the duration of the initial cluster for hold versus baseline phase. However the vowel was very clearly longer in the hold phase (by about 40ms). In the coda condition (vowel stretched, coda cluster compressed) there was no change in the vowel duration from baseline to hold phase. As in the onset condition the compressed segment was lengthened in the hold phase, but the effect was much weaker, i.e. only about 17ms. Moreover, in the onset condition a small but significant lengthening of the vowel was retained in the after-effect phase, but in the coda condition there was no evidence at all that the coda-cluster lengthening was retained in the after-effect phase.

Overall then the results indicate that subjects may be more sensitive to perturbations in



the syllable onset than the syllable-coda, but this will require confirmation with further experiments, since the perturbation scheme used here also turned out to result in greater perturbation in absolute terms in the onset condition. The fact that compensatory responses exclusively involved lengthening of compressed segments (rather than compressing of lengthened segments) suggests possible differences in the representation of temporal compared to spatial (formant) structure, for example a greater relevance of relative rather than absolute patterns in the temporal case. We are currently planning to investigate even more fine-grained temporal manipulations for evidence of adaptive responses, for example stretching C1 and compressing C2 within a cluster.

### References

- Cai, S., Ghosh, S. S., Guenther, F. H., Perkell, J. S. (2011). Focal manipulations of formant trajectories reveal a role of auditory feedback in the online control of both within-syllable and between-syllable speech timing. *J. Neurosci.* 31(45): 16483-16490.
- Goldstein, L., Pouplier, M. (2014). The temporal organization of speech, in *The Oxford Handbook of Language Production*, V. Ferreira, M. Goldrick, and M. Miozzo (eds.). Oxford: Oxford University Press
- Tourville, J., Cai, S. & Guenther, F. (2013). Exploring auditory-motor interactions in normal and disordered speech. *Proceedings of Meetings on Acoustics*, Vol. 19, 060180

## The perceptual learning of time-compressed speech: a comparison of training protocols with different levels of difficulty

**Yafit Gabay, Avi Karni, Karen Banai**

University of Haifa, Israel

*ygabay@edu.haifa.ac.il*

Although speech perception can substantially improve with practice (perceptual learning) even in adults, it is not clear how differences in the training protocol contribute to this type of learning and particularly to the ability to generalize the gains attained in training to previously encountered items (tokens) and to different talkers. Here we compared the effects of four training protocols that differed in whether and how task difficulty was changed during a training session in which participants judged the semantic plausibility of sentences presented as time compressed speech; in all other aspects the training experience was identical across conditions. In two conditions task difficulty (level of time-compression) was kept constant throughout the training session (constant-difficult, constant-easy) whereas in two other conditions task difficulty was changed in an adaptive manner (from easy to difficult, adaptive-incremental; or using a staircase procedure, adaptive-staircase). Feedback was afforded in all conditions. Compared to a control group (no training), all four protocols resulted in significant post-training gains in the ability to reproduce the sentences accurately, however, the constant-difficult protocol elicited the smallest gains. Overall, trained listeners reproduced familiar tokens presented by a new talker more accurately than untrained listeners; there was, in addition, a small advantage in reproducing novel items after training in the adaptive conditions compared to the constant difficulty conditions. The results suggest that protocols that afford an easy to difficult progression or constant easy task conditions might be advantageous for learning to decipher less familiar variations in speech; training in consistently difficult conditions is not.

Abstracts Poster

Session 2

Thursday 10:30 - 12:30 am

## Quantifying coarticulation versus invariance in German

Dzhuma Abakarova<sup>1</sup>, Khalil Iskarous<sup>2</sup>, Jan Ries<sup>1</sup>, Aude Noiray<sup>1,3</sup>

<sup>1</sup>University of Potsdam, Germany; <sup>2</sup>University of Southern California, USA; <sup>3</sup>Haskins Laboratories, USA

*dzhuma.abakarova@uni-potsdam.de*

Coarticulation, defined as varying degrees of articulatory overlap between segments, is a very important source of variability in speech. However, finding reliable methods allowing for the quantification of coarticulation versus invariance has remained an empirical challenge. Recently, Iskarous et al (2013) suggested to view coarticulation and invariance as the two ends of the scale indicating how permeable or resistant to coarticulation a segment is. The position of a segment on this scale, or the degree to which it resists coarticulation with neighboring segments is determined with Mutual Information (MI). MI measures the amount of information about segment B that is present during the production of segment A. The MI values for different aspect of segment production (e.g. various points on the tongue, lips, jaw) indicates which of them are crucial for segment production and which can vary, thus shedding light on the connection between articulatory complexity and variance in speech production.

In our study, we applied MI calculation to ultrasound data. We ask whether MI values for German consonants obtained from ultrasound data correspond to the ones found using electromagnetic midsagittal articulography (EMA) in Iskarous et al (2013) study. Although ultrasound imaging presents practical advantages in terms of data acquisition ease, the nature of data poses quantification challenges due to absence of fixed reference points. In our study, we compared MI results derived from the highest point of the tongue to those from the whole tongue as described with FDA.

### References

- Iskarous, K., Mooshammer, T., Recasens, D., Hoole, Shadle, C., Saltzman, E, and Whalen, D.H. (2013). The Coarticulation/Invariance Scale: Mutual Information as a measure of coarticulation resistance, motor synergy, and articulatory invariance, *JASA*, 134, 1271-1293.

## How different motions affect lexical access and linguistic structure in a spontaneous speech task

Susanne Fuchs<sup>1</sup>, Uwe Reichel<sup>2</sup>, Amélie Rochet-Capellan<sup>3</sup>

<sup>1</sup>ZAS, Germany; <sup>2</sup>Hungarian Academy of Sciences, Hungary; <sup>3</sup>GIPSA-lab – CNRS, France

*fuchs@zas.gwz-berlin.de*

According to dual task experiments motion plays a detrimental role for cognitive performance (Lamberg & Muratori 2012), because resources need to be shared to fulfil the two tasks. However, it has also been shown that motions are beneficial for lexical access (Krauss 1998), for associative selection of words (Oppezzo & Schwartz 2014) and for learning and memorize a new language (Schmidt-Kassow et al. 2013). We combine both views by suggesting that motion facilitates lexical access of content words, but has some drawbacks on the syntactic complexity. Moreover, we suggest that different limbs movements (legs vs. arms, right hand vs. left hand) may affect syntax and lexical access to different degrees.

Three experiments were carried out involving a spontaneous speech task: in experiment 1 we compared leg motions (biking) with different degrees of effort, in experiment 2 we studied arm versus leg motion using a minibike, and in experiment 3 we compared right versus left arm motions in right handers. In all experiments the dual tasks were also compared with single tasks. Motion, breathing and speech acoustics were recorded simultaneously. Motion and breathing frequency, syntactic complexity (number of clauses, conjunctions for main and embedded clauses) and the used vocabulary (number of content and function words; single word classes) were analysed. The results are in agreement with our predictions.

### References

- Krauss, R.M. (1998). Why do we gesture when we speak? *Current Directions in Psychological Science* 7, 54-59.
- Lamberg, E. M., & Muratori, L. M. (2012). Cell phones change the way we walk. *Gait & Posture*, 35(4), 688-690.
- Oppezzo, M., & Schwartz, D. L. (2014). Give your ideas some legs: The positive effect of walking on creative thinking. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(4), 1142-1152.
- Schmidt-Kassow, M., Deusser, M., Thiel, C., Otterbein, S., Montag, C., Reuter, M., ... & Kaiser, J. (2013). *Physical exercise during encoding improves vocabulary learning in young female adults: a neuroendocrinological study.* *PLoS one*, 8(5), e64172.

## Rate variation in speech and other motor activities

**Cécile Fougeron, Simone Falk**

Laboratoire de Phonétique et Phonologie, UMR 7018, CNRS/Paris 3 Sorbonne Nouvelle, France

*cecile.fougeron@univ-paris3.fr*

This study addresses the question of whether and how the dynamic control of speech movements differs from that of non-speech movements. This question is highly relevant in clinical phonetics where non-speech tasks are often used to access speech motor control impairments.

In a first study, a non-speech articulatory diadochokinetic task was run on 40 healthy French speakers. Non-sense homorganic CV sequences were repeated for 5 seconds as quickly as possible in a clear manner. Alternating motion rates (repetition of the same CV syllable, /bababa/) were slower than sequential motion rate (sequence of CV syllables, /badego/). Interestingly, alternating CVs showed a regular motion rate, while the /badego/ sequences have been spontaneously reinterpreted rhythmically as a French word with regular final lengthening on the last syllable /go/. This suggests that rhythmical organization of movements – typically required for fluent speech – may, by it-self, affect the temporal unwinding of motor skills.

In a second study, we will therefore explore to what extent additional factors, such as preferred tempo, the capacity to modulate rate and the meaningfulness of speech shape the dynamics of non-speech and speech motor coordination. A second aim is to find potential links between non-speech and speech movements. A pilot testing will be run with 10 French speakers. Motor skills in three different modalities will be compared: non-verbal motor (i.e., rhythmic finger tapping), articulatory motor (i.e., nonsense syllable productions; alternating and sequential, see above) and speech motor skills (i.e., production of meaningful sentences). Particular attention to speaker specific patterns will be drawn.

## Towards a typology of consonant coarticulation: gauging the space between universal and language-specific patterns of consonant timing

**Stefania Marin<sup>1</sup>, Marianne Pouplier<sup>1</sup>, Philip Hoole<sup>1</sup>, Manfred Pastötter<sup>1</sup>, Lasse Bombien<sup>1</sup>, Ioana Chitoran<sup>2</sup>, Alexei Kochetov<sup>3</sup>**

<sup>1</sup>Institute of Phonetics and Speech Processing, University of Munich, Germany; <sup>2</sup>CLILLAC-ARP, Paris Diderot University, France; <sup>3</sup>Department of Linguistics, University of Toronto, Canada

*marin@phonetik.uni-muenchen.de*

In the context of growing evidence for linguistic diversity in consonant coarticulation patterns [2-6], increasing our knowledge about the cross-linguistic range of possible coarticulatory patterns is important for understanding the relationship between physiological constraints and linguistic diversity and for understanding the degree to which coarticulation is learned. Segmental composition effects in consonant cluster timing have been attributed to physiological and perceptual factors [1, 3], but at the same time, segmentally identical consonant clusters seem to differ in degree of overlap across languages [2, 6]. Yet the extent to which this is truly the case is far from clear since direct cross-linguistic comparisons are rare. Methodological differences between experiments limit the possibility of meta-studies, especially since articulatory timing measures are highly sensitive to data treatment (e.g. filter settings). In the current study, we address this issue by employing unified data treatment and analysis procedures for articulography data from seven languages (American English, German, French, Romanian, Polish, Russian, Georgian), comparing the degree of consonant overlap across languages and clusters. We hypothesize that the impact of presumed articulatory and perceptual constraints on consonant timing will be greater for highly coarticulating languages like English in contrast to languages like Russian in which consonants coarticulate overall less. This should lead to a greater range of timing differences between clusters in the former type of languages than in the latter. The general degree of consonant coarticulation may thus also be implicated in the phonotactic cluster inventory found in a given language.

### References

- [1] Bombien, L., C. Mooshammer & P. Hoole. 2013. Articulatory coordination in word-initial clusters of German. *Journal of Phonetics*, 41, 546–561.
- [2] Bombien, L., & Hoole, P. (2013). Articulatory overlap as a function of voicing in French and German consonant clusters. *Journal of the Acoustical Society of America*, 134, 539-550.
- [3] Chitoran, I., Goldstein, L., & Byrd, D. (2002). Gestural overlap and recoverability: Articulatory evidence from Georgian. In C. Gussenhoven, T. Rietveld & N. Warner (Eds.), *Papers in Laboratory Phonology 7* (pp. 419-448). Berlin: Mouton de Gruyter.

- [4] Kochetov, A., Pouplier, M., & Son, M. (2007). Cross-language differences in overlap and assimilation patterns in Korean and Russian. *Proceedings of the XVI International Congress of Phonetic Sciences, Saarbrücken*, 1361-1364.
- [5] Pouplier, M., & Beňuš, Š. (2011). On the phonetic status of syllabic consonants: Evidence from Slovak. *Journal of Laboratory Phonology*, 2, 243-273.
- [6] Zsiga, E. 2000. Phonetic alignment constraints: Consonant overlap and palatalization in English and Russian. *Journal of Phonetics*, 28, 69-102.

## How we regulate speech rate: phonetic evidence for a 'gain strategy' in speech planning

Joe Rodd<sup>1,2</sup>, Hans Rutger Bosker<sup>1,3</sup>, Mirjam Ernestus<sup>2,1</sup>, Louis ten Bosch<sup>2</sup>, Antje S. Meyer<sup>1</sup>

<sup>1</sup>Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands;

<sup>2</sup>Centre for Language Studies, Radboud University, Nijmegen, The Netherlands;

<sup>3</sup>Donders Institute for Brain, Cognition and Behaviour, Radboud University, The Netherlands

*joe.rodd@mpi.nl*

Speech can be produced at different rates. The ability to produce faster or slower speech may be thought to result from executive control (EC) processes enlisted to modulate lexical selection and phonological encoding stages of speech planning. However, the EC strategies adopted to achieve accelerated or decelerated speech and their phonetic consequences remain unknown. This study compared two different cognitive accounts: (1) EC in the form of adjusted input activation levels to the planning network (the *gain strategy*) predicts that rate modulation is achieved by adjusting gesture duration, whilst overlap between gestures remains constant. Conversely, (2) EC in the form of adjusted selection thresholds within each planning stage (the *threshold strategy*) predicts modulated overlap and stable gesture durations across rates.

In our picture naming experiment, twelve participants were recorded (in Dutch) naming prefamiliarised (C)CV.CVC words arranged on a 'clock face' at three speaking rates (132wpm, 93wpm, 66wpm) indicated by a cursor. We measured acoustically evident gestural overlap by identifying above-average MFCC instability falling between the MAUS-aligned vowel and consonant centres. From this metric and manually corrected word onsets and offsets, we derived three dependent measures: gestural overlap duration, first syllable duration, and second syllable duration.

Mixed-effects models showed no difference in gestural overlap duration across the rate conditions. Relative to the middle rate, syllable gestures were shorter in the fast condition, and longer in the slow condition (particularly in the second, unstressed syllables). These results support an account where the *gain strategy* is the dominant mechanism of rate control, with a subordinate role for the *threshold strategy*.

## On the domain of intra-syllabic coarticulation in German children

Aude Noiray<sup>1,2</sup>, Dzhuma Abakarova<sup>1</sup>, Elina Rubertus<sup>1</sup>, Jan Ries<sup>1</sup>

<sup>1</sup>Laboratory for Oral Language Acquisition (LOLA), University of Potsdam, Germany;

<sup>2</sup>Haskins Laboratories, New Haven, USA

anoiray@uni-potsdam.de

Finding the minimal domain of language organization in the speech of children has been one of the most vivid quests in the developmental field. In the past two decades, developmental research in the domain of coarticulation has outlaid conflicting results and hypotheses as to the temporal organization of consonant-vowel syllables (CV) in the first years of language acquisition. Up to date, the question is still debated (cf. Labphon 2015).

To carry on with this endeavor, the present study reports on a cross-sectional investigation of coarticulation patterns in German children from the preschool years to the beginning of second grade as compared to adults. Unlike previous studies, we investigated the articulatory mechanisms that may account for differences in intra-syllabic coarticulation. Using ultrasound imaging, we recorded movement from the tongue during repetitions of C<sub>1</sub>VC<sub>2</sub>e nonwords varying in vowels as well as consonants' place and manner of articulation. To test whether the organization of coarticulatory patterns differed as a function of the articulatory demands for consecutive phonemes, we measured the degree of overlap versus resistance between C1 and V.

While results highlighted greater coarticulation degree for /b, g/+V compared to /d, z/+V for both children and adults, children exhibited overall greater intrasyllabic coarticulation than adults regardless of the consonantal context. As the degree of speech motor maturity may initially be circumscribed by their experience with their native language, we tested for additional effect of vocabulary and phonological development on coarticulatory patterns. The interpretation of both datasets is in progress.

### References

- Satellite *Exploring Speech Planning and Production in Children*, org. K. Demuth, S. Shattuck-Hufnagel, I. Yuen, J. Thorson, Labphon 2015, Cornell University.

## V-to-V coarticulation direction in children and adults – an ultrasound study

Elina Rubertus<sup>1</sup> Jan Ries<sup>1</sup>, Aude Noiray<sup>1,2</sup>

<sup>1</sup>Laboratory for Oral Language Acquisition (LOLA), University of Potsdam, Germany;

<sup>2</sup>Haskins Laboratories, New Haven, USA

rubertus@uni-potsdam.de

Changes in coarticulation magnitude across childhood have been interpreted as evidence for development in speech motor control, linguistic organization, or planning processes. A comparison between coarticulatory directions might help to disentangle underlying processes: While anticipatory coarticulation has been claimed to result from articulatory planning, carryover coarticulation would mainly arise because of mechanical constraints and articulator inertia (Recasens, 1987). To investigate the development of speech and coarticulatory patterns, both directions need to be taken into account.

The present study is the first to compare anticipatory and carryover V-to-V coarticulation in German children (3y, 4y, 5y, & 7y) and adults. With ultrasound imaging tongue positions were directly traced instead of inferred from the acoustic signal. A symmetrical stimulus structure (eC1VC2e) allowed us to test influences of the medial tense long vowel (/i/, /y/, /u/, /a/, /e/, /o/) on both schwas – the preceding one for anticipatory and the following one for carryover coarticulation effects. To investigate whether different mechanisms underlie the two directions, the intervocalic consonants varied in coarticulatory resistance (/d/ > /g/ > /b/). It was hypothesized that a resistant intervening consonant would decrease V-to-V coarticulation to a greater extent in the carryover (mechanical) than in the anticipatory (planning) direction (Recasens, 1987).

First results reveal that anticipatory V-to-V coarticulation is present in all cohorts. Anticipation is stronger for younger than for older participants and as expected is not affected by the resistance degree of the intervocalic consonant in any of the age cohorts. Data for carryover coarticulation are currently under analysis.

### References

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

## Exemplar theoretic modeling of phonetic convergence in dialogs

Natalie Lewandowski, Daniel Duran

Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart, Germany

natalie.lewandowski@ims.uni-stuttgart.de

As a follow-up to Lewandowski's proposal of a hybrid model of phonetic convergence (i.e. partially automatic but prone to various within- and inter-speaker influences; Lewandowski [2012]) we set out to investigate the exemplar dynamics of speaker adaptation following such a framework. Current research into speaker accommodation and variation increasingly focuses not only on the socially-relevant details of an encounter but also on speakers' individual differences – of both psychological (personality related) and cognitive (processing skill-related) nature (see Babel and McGuire, 2015, Lewandowski, 2013 and Vais et al., 2015). We propose an exemplar-theoretic model of convergence which incorporates these factors. Our computational simulation model is based on empirical data which combines dialog speech recordings with personality and cognitive data for all participants. A recently completed data collection extends the existing GECO database [Schweitzer and Lewandowski, 2013] with new dialogs from male and female subjects along with their individual personality and psychological data with a particular focus on general attention and attention to phonetic detail [Schweitzer et al., 2015]. We assume an exemplar-based speech production-perception loop which builds new productions on a collection of previously encountered speech items stored in memory. Taking into account recency, exemplars just experienced from an interlocutor may serve as or influence one's own production targets. This, however, may be enhanced or hampered by a person's personality trait combination and their respective attention skills.

### References

- M. Babel and G. McGuire. The effects of talker variability on phonetic accommodation. In *Proceedings of ICPHS 18*, 1–5, Glasgow, UK, 2015. Paper number 661.
- N. Lewandowski. *Talent in nonnative phonetic convergence*. Doctoral dissertation, Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart, 2012.
- N. Lewandowski. Phonetic convergence and individual differences in non-native dialogs. In *New Sounds*, Montreal, Canada, 2013. Presentation.
- A. Schweitzer and N. Lewandowski. Convergence of articulation rate in spontaneous speech. In *Proceedings of the Interspeech 2013, Lyon*, 525–529, 2013.
- A. Schweitzer, N. Lewandowski, and D. Duran. Attention, please! Expanding the GECO database. In *Proceedings of ICPHS 18*, Glasgow, UK, 2015. Paper number 620.
- J. Vais, M. Walsh, and N. Lewandowski. Investigating frequency of occurrence effects in L2 speakers: Talent matters. In *Proceedings of ICPHS 18*, Glasgow, UK, 2015. Paper number 723.

## Word length modulates the effect of emotional prosody

Seung Kyung Kim

Aix Marseille Univ, CNRS, LPL, Aix-en-Provence, France

seung-kyung.kim@lpl-aix.fr

Previous work on the effect of emotional prosody on spoken word recognition (Kim, 2015; Kim & Sumner, submitted) has shown that emotional prosody, independent of the lexical carrier, activates words associated with the emotional information. For example, hearing a non-emotional word (e.g., *pineapple*) uttered with angry prosody facilitates recognition of angry-associated words (e.g., *mad*). This work showed that phonetically cued social information can directly activate lexical representations.

Building on this finding, the current study delves into the nature of the affective/prosodic priming between emotional prosody and emotional words and tests if word length modulates affective priming. Word length is an important dimension in lexical processing, as longer words are shown to produce stronger lexical activation than shorter words (Pitt & Samuel, 2006). If we hypothesize that social information shows a stronger effect when lexical activation is *weaker*, we predict to find stronger affective priming with shorter words than longer words.

This hypothesis was tested with a cross-modal lexical decision task. The visual target words were 12 angry related words (e.g., *mad*, *upset*). The target was preceded by two-, three-, or four-syllable non-emotional prime words (e.g., *atom*, *envelope*, *aluminum*) spoken with angry prosody. The results showed that listeners recognized angry words faster after hearing angry prosody than after hearing neutral prosody when the prime words were short (2 syllables) but not when the prime words were longer (3–4 syllables). The current results provide evidence that social effects in word recognition are modulated by the strength of lexical activation.

## Highly complex syllable structure: a motivated and stable feature

**Shelece Easterday**, Department of Linguistics, University of New Mexico, USA  
seasterd@unm.edu

I present typological evidence that highly complex syllable structures (HCSSs) are stable structures, despite their rarity and the challenges they pose to phonological theory. HCSSs – e.g., *O'odham* *ʔaŋspk* (Hill & Zepeda 1992) — are rare, both cross linguistically and in type frequency within languages. Overwhelmingly, languages with HCSSs permit long strings of obstruents, often voiceless. These extreme deviations from the idealized CV type are problematic for abstract theoretical treatments of the syllable, being unmotivated in physiological, phonological, and acquisition accounts. This raises questions as to how and why these structures arise and persist in speaker populations.

A common characteristic of large tautosyllabic obstruent clusters is the presence of open transitions between the consonants. The resulting perceptual effect is alternately described as aspiration, a short voiceless vowel, or a brief transitional vocoid. These transitions facilitate both the production of such sequences and their perception, providing the listener with crucial acoustic cues. Comparative, historical, and instrumental evidence suggest that rampant vowel reduction is often responsible for the development of these clusters and their characteristic phonetic properties (cf. Chitoran & Babliyeva 2007 for Lezgian).

Analysis of a diversified sample of 24 languages with HCSSs reveals that the articulatory routines contributing to the emergence of these structures may persist over long periods of time. Roughly half of these languages have active processes of vowel deletion, many of which result in the formation of canonical or non-canonical tautosyllabic clusters of obstruents. By comparison, very few languages have processes of vowel epenthesis or cluster simplification targeting obstruent clusters. For example, where epenthesis occurs in the language sample, it tends to target consonant sequences with sonorants. Consonant cluster simplification processes occur in the sample, but are most frequent and regular in languages where HCSSs are a marginal pattern. The results here indicate that despite theoretical issues of analysis, HCSSs are neither problematic for speakers nor unstable in speech communities. The phonetic processes responsible for creating these syllable patterns appear to be remarkably persistent, and more prevalent than processes which 'repair' these structures. I suggest that the persistence of these rare structures is at least partly motivated and maintained through the salient phonetic properties of obstruent clusters, which accommodate both the speaker and hearer.

### References

- Chitoran, Ioana and Ayten Babliyeva. 2007. An acoustic description of high vowel syncope in Lezgian. *International Congress of Phonetic Sciences XVI*: 2153-56.
- Hill, Jane H. and Ofelia Zepeda. 1992. Derived words in Tohono O'odham. *International Journal of American Linguistics* 58(4): 355-404.

## Phonetic variation and contrast neutralization patterns in Romanian fricatives across different speech styles

**Laura Spinu<sup>1</sup>, Ioana Vasilescu<sup>2</sup>, Bianca Vieru<sup>3</sup>, Lori Lamel<sup>2</sup>, Jason Lilley<sup>4</sup>**

<sup>1</sup>Univ. of Toronto, Canada; <sup>2</sup>LIMSI-CNRS, France; <sup>3</sup>Vocapia Research, France; <sup>4</sup>Nemours Biomedical Research, USA

lspinu@udel.edu

We compare the acoustic properties of Romanian fricatives across three different speech styles: controlled experimental speech (n=31) [5], (b) continuous speech from news broadcasts (141 speakers) [3, 6] and (c) conversational speech collected using the Map Task approach [following 1, 2]. Previous work with Romanian controlled speech [4, 5] identified two phenomena related to fricatives: (1) /h/ is realized as glottal, velar and palatal, with a complex allophonic/free variation distribution, and (2) the plain-palatalized contrast (which has high functional load as a number and person marker) is partially neutralized with alveolars and postalveolars. We use a novel tool [5] to identify the properties of these fricatives and neighboring vowels. The feature set employed for coding their acoustic properties comprises cepstral coefficients 0-5. We measure the acoustic features at 10-ms intervals; the measures obtained are then binned into three contiguous regions for each segment (onset, steady state, and offset). The boundaries between regions are set using a hidden Markov model to determine three internally uniform regions with respect to their acoustic properties. The mean value of each acoustic feature within each region is then obtained. This enables us to identify changes over time, which are crucial in the case of partial neutralization. By identifying (1) the various acoustic manifestations of /h/ and the frequency/environment of their occurrence, and (2) the neutralization patterns of the plain-palatalized contrast, this study documents the properties of Romanian fricatives and contributes to our understanding of the dynamics of sound change in phonological systems.

### References

- [1] Anderson, A., Bader, M., Bard, E., Boyle, E., Doherty, G. M., Garrod, S., et al. 1991. The HCRC map task corpus. *Language and Speech*, 34, 351-366.
- [2] Baker, R. and Hazan, V. 2011. DiapixUK: task materials for the elicitation of multiple spontaneous speech dialogs. *Behavior Research Methods* 43:761-770.
- [3] Renwick, MEL, Vasilescu, I., Dutrey, C., Lamel, L., Vieru, B. 2016. Marginal Contrast Among Romanian Vowels: Evidence from ASR and Functional Load. *Interspeech 2016*, 2433-2437.
- [4] Spinu, L., Vogel, I., and Bunnell, T. (2012). Palatalization in Romanian: Acoustic properties and perception. *Journal of Phonetics* 40(1), 54-66.
- [5] Spinu, L. and Lilley, J. 2016. A comparison of cepstral coefficients and spectral moments in the classification of Romanian fricatives. *Journal of Phonetics* 57, 40-58.
- [6] Vasilescu, I., Vieru, B., Lamel, L. 2014. Exploring pronunciation variants for Romanian speech-to-text transcription. *Proceedings of the 4th Workshop on Spoken Language Technologies for Under-resourced languages (SLTU'14)* St. Petersburg, Russia.

## Comparing the rhythmic properties of plain and Lombard speech

Hans Rutger Bosker<sup>1,2</sup>, Martin Cooke<sup>3,4</sup>

<sup>1</sup>Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands, <sup>2</sup>Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands; <sup>3</sup>University of the Basque Country, Universidad del País Vasco/EHU, Spain; <sup>4</sup>Ikerbasque (Basque Science Foundation), Bilbao, Spain

HansRutger.Bosker@mpi.nl

Speech is an inherently rhythmic signal. Even though speech is by no means strictly periodic, energy patterns in speech are constrained by the physiological dynamics of the lips, jaw, and tongue. As such, energy fluctuations in speech typically occur within the 2-20 Hz range.

This rhythmicity in speech has been suggested to play a central role in comprehension, facilitating the processing of the signal and aiding recognition memory. This study investigated whether speakers actually produce more rhythmic speech (i.e., more regular alternations between high and low amplitude intervals) in acoustically challenging conditions (e.g., background noise), thus facilitating comprehension.

Four different corpora were analyzed (with varying sample sizes and talkers), each including plain speech (sentences produced in quiet) and matched Lombard speech (same sentences produced in noise). Each sentence was first normalized in amplitude by RMS scaling, thus avoiding intensity confounds. The envelope of the normalized signal was then submitted to a Fast Fourier Transform (FFT), resulting in the modulation spectrum of that particular sentence. Comparing the average modulation spectra of plain and Lombard speech revealed greater power in Lombard speech in the *delta* band (1-3 Hz), across all four corpora, which could not be attributed to decreases in speech rate.

These findings suggest that speakers produce more rhythmic speech, particularly in the 1-3 Hz range, when talking in noise (vs. in quiet). Results are discussed in terms of the functional role of rhythmicity in dialogue and potential underlying neurocognitive mechanisms (e.g., neural oscillatory dynamics).

## Phonetic cue enhancement in hyperarticulation of Korean sibilants

Jessamyn Schertz<sup>1,2</sup>, Yoonjung Kang<sup>3,2</sup>, Sungwoo Han<sup>4</sup>

<sup>1</sup>University of Toronto Mississauga, Canada; <sup>2</sup>University of Toronto, Canada; <sup>3</sup>University of Toronto Scarborough, Canada, <sup>4</sup>Inha University, South Korea

jessamyn.schertz@utoronto.ca

This work explores enhancement of two phonetic cues, VOT and f0, in the hyper articulation of Korean sibilants. Korean has a two-way fricative contrast, fortis /s/ and “nonfortis” /s/, which is ambiguous with respect to the three-way laryngeal contrast characterizing stops and affricates. We examine enhancement of two contrasts involving /s/: /s/ vs. /ss/, and /s/ vs. lenis affricate /c/.

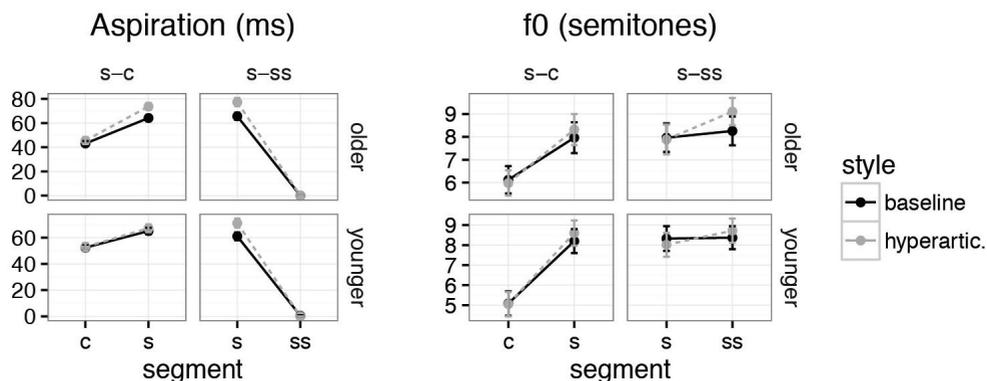
We analyzed use of f0 and aspiration in baseline and hyper articulated productions by younger (n=33, aged 19-53) and older (n=29, aged 54-82) Seoul Korean speakers. Both cues differentiated the /s/-/c/ contrast in baseline productions, but their relative importance is in flux: younger speakers used more f0, and less aspiration, than older speakers. Production of /s/-/ss/ was relatively stable across ages, differing in aspiration but not f0.

In hyperarticulation, we expected speakers to enhance dimensions as a function of their relative importance in baseline productions; however, this was not always the case (italic text in Table 1). Most notably, hyperarticulation led to “emergence” of an f0 distinction in the /s/-/ss/ contrast, indicating that despite a lack of f0 distinction in baseline, the contrast is, at some level, characterized by higher f0 for fortis /ss/, in line with previous *perceptual* findings. Follow-up comparisons showed that speakers’ use of f0 in production correlated with their reliance on f0 in a perception task, providing evidence for an individual-level perception-production link in phonetic representation that is only visible when considering the hyperarticulated productions. We frame the results within the broader context of the ongoing sound change in the Korean laryngeal system.

		/s/ vs. /c/		/s/ vs. /ss/	
		Baseline contrast?	Enhanced?	Baseline contrast?	Enhanced?
f0	Older	yes: s > c	yes	<i>no</i>	<i>yes (ss &gt; s)</i>
	Younger	yes: s > c	yes	<i>no</i>	<i>yes (ss &gt; s)</i>
	Change over time?	Increase		No change	
aspiration	Older	yes: s > c	yes	yes: s > ss	yes
	Younger	<i>yes: s &gt; c</i>	<i>no</i>	yes: s > ss	yes
	Change over time?	Decrease		No change	

**Table 1:** Baseline and enhanced use of f0 and aspiration in differentiating two phonetic contrasts (/s/ vs. /c/ and /s/ vs. /ss/) by older and younger Seoul Korean speakers. →

Th



**Figure 1:** Mean production values for aspiration (in ms) and f0 (in semitones) across contrasts, speech styles, and age groups. Error bars show one standard error based on by-speaker means.

## Temporal organization of initial fricative clusters in simultaneous bilingual speakers

Conceição Cunha, Lorenzo Spreafico

Free University of Bozen-Bolzano, Italy

[cunha@phonetik.uni-muenchen.de](mailto:cunha@phonetik.uni-muenchen.de)

The main aim of this paper is to investigate whether bilingual speakers use two different timing patterns for the same sequence of consonants as has been showed for monolingual speakers of the two languages they speak. Indeed, articulatory studies attested simplex timing of consonants in /sC(C)/ clusters (Hermes, 2013), while Tyrolean, in the light of its genetic and geographical proximity to German, is expected to present similar complex timing of consonants in sibilant clusters (Pouplier, 2012). How bilingual speakers equally exposed to both languages since birth deal with these timing differences is completely unclear. This will be the main concern of the paper.

Unlike previous research, we examine timing patterns and possible physical correlates of syllabic organization in Italian and Tyrolean by means of the ultrasound tongue imaging technique (UTI). Data are collected using an Ultrasonix Tablet system (100 Hz, FOV 148°). Five female bilingual speakers born and living in the South Tyrolean Underland are asked to read real words three times from a randomized list. The list includes real Italian (n= 66) and Tyrolan (n= 93) words containing /C(C)/ and /sC(C)/ clusters in word initial position. Tongue contours are extracted from raw US images using the AAA software (Articulate Instruments, 2012). To extract dynamic information, gestural landmarks are identified using the PD-IWT algorithm (Vietti et al., in preparation). First results attested greater differences between /C(C)/ and /sC(C)/ sequences in Tyrolean than in Italian.

### References

- Articulate Instruments Ltd 2012. Articulate Assistant Advanced User Guide: Version 2.14. Edinburgh, UK: Articulate Instruments Ltd.
- Hermes, A., Mücke, D. & M. Grice 2013. Gestural coordination of Italian word initial clusters – The case of ‘impure s’. *Phonology*. 30(1).
- Pouplier, M. 2012. The gestural approach to syllable structure: Universal, language- and cluster specific aspects. In Fuchs, Weihrich, Pape, Perrier (ed.s) *Speech Planning and Dynamics*. Peter Lang. 63-96.

## Individual acoustic cues to distinctive features: an approach to quantifying and categorizing large samples of dynamic speech data

Stefanie Shattuck-Hufnagel, Jeung-Yoon Choi, Rebeka Bell

Speech Communication Group, Massachusetts Institute of Technology, USA

sshuf@mit.edu

Recent decades have seen the emergence of substantial evidence that speakers and listeners represent, control and make use of systematic patterns of subfeatural variation in the surface phonetic forms of words and their sounds. Quantifying this phonetic variability is challenging, because many dimensions are changing at once and many contributing factors have been identified. One promising metric is based on the modification of individual acoustic cues to distinctive features, following up on the proposal by Stevens (2002) that individual feature cues are the focus of the initial steps in human speech perception. Drawing on the work of Halle (1995), Stevens proposes that certain aspects of the speech signal provide particularly useful information for identifying the distinctive features of the speaker's intended words, and thus for accessing the appropriate lexical items. He identifies two major types of feature cues, i.e. *acoustic landmarks* (abrupt spectral changes which signal manner features such as [consonantal], [vocalic], [obstruent] etc.) and *other acoustic cues* (acoustic patterns, often found in the vicinity of landmarks, which signal place and voicing features), as well as other acoustic events (such as the onset and offset of pulses signaling vocal fold vibration; irregular pitch periods; and formant changes resulting from the opening and closing of the velopharyngeal port). We will present results from hand-labelling of individual feature cues in read and spontaneous speech, and from algorithms for automatic detection of these cues, providing quantitative measures of this aspect of systematic context-governed surface phonetic variation.

### References

- Halle, M. (1992), Phonological features. In W. Bright (ed.), *International encyclopedia of linguistics* Vol. 3. 207– 212. Oxford: Oxford University Press.
- Stevens, K.N. (2002), Toward a model for lexical access based on acoustic landmarks and distinctive features. *The Journal of the Acoustical Society of America* 111 (4). 1872- 1891.

## The interplay of phonetics and phonology in speech development – a study of children with dysarthria

Theresa Schölderle, Elisabet Haas, Wolfram Ziegler

Clinical Neuropsychology Research Group, Institute for Phonetics and Speech Processing, Ludwig Maximilians University, Munich, Germany

theresa.schoelderle@ekn-muenchen.de

The formation of phonological representations requires abstraction from phonetic diversity (e.g., recognizing a word spoken by different speakers as the same). In typically developing children this process of *phonologization* interacts with the continuing increase of articulatory proficiency as the children's own vocal productions are successively integrated into the emerging phonological concepts (1, 2). This raises the question of how phonological representations emerge in children with early acquired, persisting disorders of phonetic processes, such as in *dysarthria* (neurologic speech disorder) (3). The current study investigates (a) if childhood dysarthria interferes with the process of phonologization and (b) to what extent characteristics of dysarthria can be mapped onto aspects of phonological processing. A minimal pair discrimination task was developed. To assess phonological abstraction, the stimuli pertaining to each target word were phonetically diverse (different speakers; manipulation of acoustic properties). We chose minimal pairs with contrasts that are typically reduced in the speech of children with dysarthria (e.g., *oral/nasal* in children with *hypernasality*). The experiment is administered to dysarthric as well as typically developing children of pre-school age.

Examinations and analyses are still under way. We expect to find differences between dysarthric and typically developing children regarding their stage of phonologization (e.g., a higher sensitivity towards phonetic diversity in dysarthric children) that can be related to the characteristics of dysarthria in the neurologically impaired group. Our findings will have implications for current models of language acquisition as well as for clinical assessment of children with dysarthria.

### References

- 1. Vihman MM. Phonological Development: The First Two Years. Oxford: Wiley Blackwell; 2014.
- 2. Vihman MM. Ontogeny of Phonetic Gestures: Speech Production. In: Mattingly IG, Studdert-Kennedy M, editors. Modularity and the Motor Theory of Speech Perception Proceedings of a Conference to Honor Alvin M Liberman. New York: Lawrence Erlbaum Associates; 1991. p. 69-84.
- 3. Peeters M, Verhoeven L, de Moor J, Van Balkom H. Importance of speech production for phonological awareness and word decoding: The case of children with cerebral palsy. *Research in developmental disabilities*. 2009;30(4):712-26.

## Do patients with speech production impairments benefit from auditory priming with a regular metrical pattern?

Ingrid Aichert, Katharina Lehner, Simone Falk, Mona Späth, Wolfram Ziegler

EKN, Institute of Phonetics and Speech Processing, Ludwig Maximilians University, München, Germany

ingrid.aichert@ekn-muenchen.de

**Background** – In earlier investigations based on repetition tasks we found that the regular metrical pattern in German (trochee), unlike the iambic pattern, facilitates articulation in patients with apraxia of speech (e.g. Aichert et al. 2016), confirming that segmental and prosodic aspects of speech production interact. In this presentation, we investigated if articulation in apraxic speakers also benefits from auditory priming by regular metrical patterns.

**Method** – 12 patients with apraxia of speech, 12 aphasic patients with phonological impairment and 36 neurologically healthy speakers were examined. A sequential synchronization paradigm based on a sentence completion task was conducted in four different conditions:

- (1) Regular prime sentence, trochaic target word
- (2) Regular prime sentence, iambic target word
- (3) Irregular prime sentence, trochaic target word
- (4) Irregular prime sentence, iambic target word.

The participants' responses were phonetically transcribed. Speech errors were classified by error types (e.g. segmental errors, prosodic errors, groping). Furthermore, responses were analyzed for onset latencies of the target words and for metrical parameters.

**Results** – Data evaluation is still in progress.

**Discussion** – Results will be interpreted on the background of models of speech production and apraxia of speech and will be further discussed with reference to rhythmic synchronization accounts of between-speaker interaction (e.g., Cummins, 2009).

### References

- Aichert, I., Späth, M., Ziegler, W. (2016). The Role of Metrical Information in Apraxia of Speech. Perceptual and Acoustic Analyses of Word Stress. *Neuropsychologia*, 82, 171-178.
- Cummins, F. (2009). Rhythm as entrainment: The case of synchronous speech. *Journal of Phonetics*, 37, 16-28.

## Rhotic metathesis in CVrC and CrVC syllables: evidence from Modern Croatian

Valentin Hofmann, Daniela Müller

Ludwig Maximilians University Munich, Germany

valentin.hofmann@campus.lmu.de

Common Slavic CVrC and CrVC syllables are known for their intricate diachronic developments in the individual Slavic daughter languages. For instance, as described by Shevelov (1964), Common Slavic *vīrxŭ* 'tip' (CVrC) corresponds to Polish *wierzch* (CVrC), Bulgarian *vrăx* (CrVC), Bosnian-Croatian-Serbian *vrh* (CṛC) and Northern Russian *verëx* (CVrVC). We will argue that these forms represent different degrees of rhotic metathesis and that rhotic metathesis is due to the phonetic structure of the alveolar tap [r̥] and trill [r̥̥].

As shown by Proctor (2009) and Savu (2012), the rhotics [r̥] and [r̥̥] contain two vocoids on each side of the constriction, yielding the structure 'vocoid – constriction – vocoid' (<sub>v</sub>r̥<sub>v</sub>). These vocalic elements most saliently appear adjacent to a consonant, that is, in the clusters C<sub>v</sub>r̥<sub>v</sub>V, V<sub>(v)</sub>r̥<sub>v</sub>C and C<sub>v</sub>r̥<sub>v</sub>C. They can change into full vowels if they are misinterpreted by the listener as part of the intended pronunciation. Metathesis occurs, then, if the original full vowel is simultaneously parsed by the listener as the vocoid inherent in the rhotic (Czaplicki 2013).

This theoretical account has never been subject to systematic experimental investigation. We conducted a perception experiment with native speakers of Croatian, a language that has syllabic rhotics in addition to the more common consonantal rhotics. In this experiment, we examine the qualitative and quantitative thresholds for assigning phonetic clusters of the type CV<sub>(v)</sub>r̥<sub>v</sub>C, C<sub>v</sub>r̥<sub>v</sub>VC and C<sub>v</sub>r̥<sub>v</sub>C to the phonological clusters CVrC, CrVC, CṛC and CVrVC, where the vocalic elements vary in their formant structure from vocoid to full vowels as well as in their duration from values appropriate for vocoids to values typical of long vowels. The results of our study will prove profitable for linguistic research not only with regard to rhotic metathesis, but to the understanding of long-distance metathesis in general.

### References

- Czaplicki, B. (2013). R-metathesis in English: An account based on perception and frequency of use. *Lingua* 137. 172-192.
- Savu, C. F. (2012). An acoustic-phonetic perspective on the phonological behavior of the rhotic tap. *Proceedings of ConSOLE XX*. 213-227.
- Proctor, M. (2009). *Gestural Characterization of a Phonological Class: the Liquids*. New Haven: Yale University.
- Shevelov, G. (1964). *A Prehistory of Slavic. The Historical Phonology of Common Slavic*. Heidelberg: Carl Winter.

## Effect of phrasal accent on the articulation of Slovak /l/ in different syllable positions

Lia Saki Bučar Shigemori, Marianne Pouplier

Institute of Phonetics and Speech Processing, Ludwig Maximilians University Munich, Germany

[lia@phonetik.uni-muenchen.de](mailto:lia@phonetik.uni-muenchen.de)

This study examines the effect of phrasal accent on the articulation of /l/ in the syllable nucleus, onset and coda position using ultrasound. Vowels have been argued to be the main carriers of prosodic emphasis. Under accentuation, the sonority of the vowel is enhanced by lowering the jaw and tongue and widening the oral cavity (Beckman et al., 1992). Based on Slovak, we investigate in the current study whether phrasal accent has a different effect on /l/, depending on its syllable position. Slovak has a dark /l/, which can be syllabic in both stressed and unstressed position (Král', 2005). A dark /l/ has been described as having both, a consonantal (tongue tip) and a vocalic (tongue back) gesture (Sproat and Fujimura, 1993). We want to examine whether accent differentially targets the tongue back gesture of /l/ in the nucleus position, thus strengthening the vocalic characteristics of the syllabic consonant. If /l/ is more vowel-like in the nuclear position, an enhancement of the vocalic tongue back gesture and a weakening of the consonantal tongue tip gesture is expected under accentuation.

Six native speakers of Slovak have been recorded using ultrasound. They read the pseudo words *plpap*, *plepap* and *pelpap*, embedded in two different carrier phrases to elicit two accentuation patterns. We compared the tongue shapes at the temporal midpoint of the /l/ for the two accent conditions separately for each syllable position. Preliminary results show the same pattern for accentuation in the three positions. The consonantal as well as vocalic gesture of /l/ are strengthened in the same way regardless of its syllable position, suggesting that /l/ is not more vocalic in the nuclear position.

### References

- Beckman, M. E., Edwards, J., and Fletcher, J. (1992). Prosodic structure and tempo in a sonority model of articulatory dynamics. *Papers in Laboratory Phonology II*, pages 68-86.
- Král', A. (2005). *Pravidlá slovenskej výslovnosti: systematika a ortoepický slovník*. Matica slovenská.
- Sproat, R. and Fujimura, O. (1993). Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics*, 21(3): 291-311.

## Author Index

Abakarova, Dzhuma	58	Gabay, Yafit	56	Nguyen, Noël	44
Aichert, Ingrid	16, 76	Galatà, Vincenzo	11	Noiray, Aude	48, 58, 65
Asano, Yuki	51	Grohe, Ann-Kathrin	36		
Athanasopoulou, Angeliki	34			Pacchiarotti, Sara	31
		Haas, Elisabet	75	Pastätter, Manfred	61
Babel, Molly	10	Han, Sungwoo	71	Peng, Gang	20
Banai, Karen	49, 56	Hanulíková, Adriana	21	Pouplier, Marianne	47, 61, 78
Barón, Leonardo	44	Hermes, Anne	17		
Bell, Rebeka	74	Hofmann, Valentin	77		
Bombien, Lasse	61	Hoole, Philip	47, 54, 61		
Bosker, Hans Rutger	19, 33, 63, 70			Redford, Melissa	31
Broersma, Mirjam	50	Iskarous, Khalil	58	Reichel, Uwe	59
Bruni, Jagoda	35	Jakob, Hanna	16	Reinisch, Eva	38, 39
Bučar Shigemori, Lia Saki	78			Ries, Jan	48, 58, 64, 65
				Rochet-Capellan, Amélie	59
Chen, Si	37	Kaland, Constantijn	11	Rodd, Joe	63
Chitoran, Ioana	47, 61	Kamiyama, Takeki	47	Rubertus, Elina	65
Choi, Jueng-Yoon	74	Kang, Yoonjung	71		
Cooke, Martin	70	Karni, Avi	56	Schertz, Jessamyn	71
Corley, Martin	30	Kell, Christian	26	Schölderle, Theresa	75
Cunha, Conceição	73	Knight, Sara	30	Senior, Brianne	10
		Kochetov, Alexei	13, 61	Shattuck-Hufnagel, Stefanie	15
		Konieczny, Lars	21	Sichlinger, Laura	54
D'Imperio, Mariapaola	15	Krüger, Stella	48	Sjerps, Matthias	20
Dogil, Grzegorz	35	Kwon, Harim	47	Smith, Rachel	52
Dohen, Marion	15	Kyung, Seung	67	Späth, Mona	76
Drake, Eleanor	30			Spinu, Laura	13, 34, 69
Dufour, Sophie	40	Lamel, Lori	69	Spreafico, Lorenzo	11, 73
Duran, Daniel	35, 66	Lavie, Limor	49	Stuart-Smith, Jane	52
		Lehner, Katharina	16, 76		
Easterday, Shelece	68	Lennon, Robert	52	ten Bosch, Louis	63
Eger, Nikola Anna	38	Lentz, Tomas	47	Thies, Tabea	17
Egurtzegi, Ander	45	Levy, Helena	21		
Ernestus, Mirjam	63	Lewandowski, Natalie	66	Vasilescu, Ioana	69
Esteve-Gibert, Núria	15	Lilley, Jason	34, 69	Vieru, Bianca	69
		Llompart, Miquel	39	Vietti, Alessandro	11
		Loevenbruck, Hélène	15		
Falk, Simone	60, 76			Wayland, Ratreë	37
Fanny, Guitard-Ivent	43	Marin, Stefania	61	Weber, Andrea	36
Flögel, Mareike	26	Maslowski, Merel	19	Wong, Karina	10
Fougeron, Cécile	43, 60	Meyer, Antje S.	19, 63		
Fuchs, Susanne	59	Meynadier, Yohann	40		
		Mitterer, Holger	42	Zhang, Caicai	20
		Muecke, Doris	17	Zhu, Yiqing	37
		Müller, Daniela	77	Ziegler, Wolfram	16, 75, 76

## Program at a Glance

### Wednesday May 3 2017

09:20	<b>Welcome</b>
09:30 - 10:15	James McQueen
10:15 - 10:35	Molly Babel, Brianne Senior, Karina Wong
10:35 - 10:55	Alessandro Vietti, Lorenzo Spreafico, Vincenzo Galatà, Constantijn Kaland
10:55 - 11:15	<i>COFFEE BREAK</i>
11:15 - 12:00	Janet Pierrehumbert
12:00 - 12:20	Laura Spinu
12:20 - 02:00	<i>LUNCH BREAK</i>
02:00 - 02:45	Jan Edwards
02:45 - 03:05	Núria Esteve-Gibert, H�el�ene Loevenbruck, Marion Dohen, Mariapaola D'Imperio
03:05 - 03:30	<i>COFFEE BREAK</i>
03:30 - 05:00	<b>Poster Session 1</b>
05:00 - 05:20	Wolfram Ziegler, Hanna Jakob, Katharina Lehner, Ingrid Aichert
05:20 - 05:40	Doris Muecke, Anne Hermes and Tabea Thies
05:40	<b>WELCOME RECEPTION</b>
06:30	<i>DINNER</i>

### Thursday May 4 2017

08:30 - 09:15	Esther Janse
09:15 - 09:35	Merel Maslowski, Antje S. Meyer, Hans Rutger Bosker
09:35 - 09:55	Matthias Sjerps, Caicai Zhang, Gang Peng
09:55 - 10:15	Helena Levy, Lars Konieczny Adriana, Hanul�kov�a
10:15 - 10:30	<i>COFFEE BREAK</i>
10:30 - 12:30	<b>Poster Session 2</b>
12:30 - 02:00	<i>LUNCH BREAK</i>
02:00 - 02:45	Pascal Perrier
02:45 - 03:30	Adamantios Gafos
03:30 - 04:00	<i>COFFEE BREAK</i>
04:00 - 04:45	Douglas Shiller
04:45 - 05:30	Caroline Niziolek
05:30 - 05:50	Mareike Fl�gel and Christian Kell
06:30	<i>DINNER</i>

### Friday May 5 2017

08:30 - 09:15	Jennifer Cole
09:15 - 10:00	Ann Bradlow
10:00 - 10:30	<i>COFFEE BREAK</i>
10:30 - 11:15	Matthew Goldrick
11:15 - 11:35	Martin Corley, Eleanor Drake, Sara Knight
11:35 - 11:55	Melissa Redford, Sara Pacchiarotti
11:55 - 12:15	<b>Farewell, Closing Remarks</b>
12:30	<i>OPTIONAL LUNCH (IF BOOKED WITH REGISTRATION)</i>



**Research Focus**  
Language: Birth and Decay