Proceedings

4th Summerschool on

Speech Production and Perception:

Speaker-Specific Behaviour

Program

Monday, 30th September

14.30 - 16.00	**Rachel Smith (University of Glasgow): Speech Perception and Fine Phonetic Detail**	8
16.00 - 16.30	Coffee break	
16.30 - 18.00	**Adaptation and Fine Phonetic Detail** 1. Natalie Lewandowski & Antje Schweitzer "Cognitive and psychological influences on phonetic adaptation in dialogue"	10
	2. Maëva Garnier, Nathalie Henrich, Lucie Ménard, Gabrielle Richard "Inter-individual variability in speech adaptation to noisy environments"	11
	3. Marija Tabain "Effect of word position on stop bursts in Pitjantjatjara"	13
18.30	Dinner	

Tuesday, 1st October

9.00 - 10.30	**Frank Eisner (MPI, Nijmegen): Speech Perception and Listener Adjustments**	15

10.30 - 11.00	Coffee break	
	Prosody in Production and Perception	
11.00 - 12.30	1. Francesco Cangemi "Listener-specific perception of speaker-specific productions? Evidence from intonation and supralaryngeal articulation across focus structures in German"	16
	2. Iris Chuoying Ouyang & Elsi Kaiser "Individual-specific prosodic encoding and decoding of informativity"	18
12.30 - 14.30	Lunch	
14.30 - 16.00	**Benjamin Swets (Grand Valley State University, Allendale): Psycholinguistics and Planning**	20
16.00 - 16.30	Coffee break	
	Poster Session	
16.30 - 18.00	1. Cornelia Heyde "Articulatory Motor control in adults who stutter. An electromagnetic articulatory study"	22
	2. Jeanin Jügler "Feedback methods to improve phonetic and phonological skills in foreign language acquisition"	23
	3. Claudia Canevari "Articulatory normalization via imitation strategy in phone classification task"	25
	4. Nathalie Henrich & Maëva Garnier "The impact of source-filter interaction on speaker's formants and pitch variability"	26

	15. Sven Grawunder & Lucía Golluscio "Language or Speaker? – Investigating coronal-velar alternations in Vilela"	
18.30	Dinner	

Wednesday, 2nd October

9.00 - 10.30	**Alejandrina Cristia (MPI, Nijmegen): Speech Acquisition**	46
10.30 - 11.00	Coffee break	
11.00 - 12.30	**L1 and L2 Acquisition**	
	1. Alex de Carvalho; Isabelle Dautriche; Anne Christophe "Children's use of prosody to compute syntactic structure on-line"	47
	2. Sven Grawunder "Speaker and areal specific paths of acquiring neutralization of final stops in German"	49
	3. Marie-José Kolly, Volker Dellwo, Adrian Leemann "The influence of second language on speaker-idiosyncratic temporal patterns"	52
12.30 - 14.30	Lunch	
14.30 - 18.00	**Social Event**	
18.30	Dinner	

Thursday, 3rd October

9.00 - 10.30	**Pascal Perrier (GIPSA-lab, Grenoble): Biomechanics**	54
10.30 - 11.00	Coffee break	
11.00 - 12.30	**Articulation**	
	1. Guillaume Barbier, Pascal Perrier, Lucie Ménard, Yohan Payan, Mark K. Tiede, Joseph S. Perkell "Speaker-specific behavior of 4-year-old Canadian French children: ultrasound data analysis in the light of biomechanical knowledge"	56
	2. Jana Brunner, Christian Geng & Adamantios Gafos "Contributions of the vocal tract shape to the C-center effect"	58
	3. Leonardo Lancia & Sven Grawunder "New methods to explore the complexity of the behavior of the larynx and its interaction with breathing factors and configuration of the tongue"	60
12.30 - 14.30	Lunch	
14.30 - 16.00	**Melanie Weirich (Friedrich-Schiller University, Jena): Articulation**	62
16.00 - 16.30	Coffee break	

	Physiology and Speech	
16.30 - 18.00	1. Amélie Rochet-Capellan & Susanne Fuchs "Speaker-specific breathing profiles during spontaneous speech"	64
	2. Camille Fauth, Béatrice Vaxelaire, Jean-François Rodier, Pierre-Philippe Volkmar, Marion Béchet, Rudolph Sock "An acoustic study of sustained vowels produced by patients with or without recurrent laryngeal nerve paralysis after thyroid Surgery"	65
18.30	Dinner	

Friday, 4th October

9.00 - 10.30	**Roger Mundry (MPI, Leipzig):** **Tutorial Statistics**	67
10.30 - 11.00	Coffee break	
11.00 - 12.30	**Roger Mundry (MPI, Leipzig):** **Tutorial Statistics (continuation)**	

Perception of speaker-specific fine phonetic detail

Rachel Smith (University of Glasgow, UK)

As a source of systematic variation in the speech signal, "the speaker" has attracted growing interest since models of speech perception began, in the 1990s, to treat variation as information rather than noise. Early work by Pisoni, Goldinger, Nygaard and colleagues showed that perception can be facilitated when conditions allow information about the speaking voice to be encoded and accessed. This work emphasised global characteristics like f0, vocal effort and rate. In the 2000s, the focus shifted to how listeners adapt their representations of (usually segmental) phonetic categories, guided by word meaning, to accommodate speakers' idiosyncratic pronunciations of individual sounds [1].

A speaker, however, is more than a collection of phoneme realisations and some average prosodic properties. For example, speakers vary in their "prosodic signatures", i.e. the detailed phonetic means they use to index boundaries between units, and prominence of units. Long-term articulatory settings also impart characteristic "voice qualities" [2] that colour vocal output systematically, but interact in complex ways with the segments and prosody of the message. In this talk I review evidence from my own work and others' that processing of the linguistic message is affected by inter-speaker variation in a number of aspects of phonetic detail, including idiosyncratic realisation of subphonemic features (such as voicing, e.g. [3]), syllable and word boundaries [4], stress-conditioned allophony [5] and realisation of specific grammatical morphemes [6]. Following Polysp [7], these data are interpreted as suggesting that people may learn about speaker-specific realisations of any type of linguistic structure, from sub-phonemic features up to larger prosodic structures and, potentially, conversational units such as speaking turns. Speaker-specific attributes may even, on a more associative basis, enable direct access to aspects of meaning (e.g. [8]). Learning of speaker-specific patterns is hypothesized to be guided by not only lexical meaning but also phrasal and pragmatic meaning. Nevertheless, there are also circumstances under which speaker-specific phonetic detail seems not to be stored, despite relevant exposure. I discuss these counterexamples considering issues such as the frequency and salience of particular speaker-specific patterns in the input, and listener biases in attribution of variation to particular causes (e.g. [5], [9]).

Finally I attempt to build on Polysp to consider the nature of representations within the "speaker space". I propose that if listeners are to make full use of the richness of speaker information available in the signal, they need to track both "sub-speaker" variability (e.g. consequences of the speaker's temporary physical state, situational requirements, and conversational goals) as well as "supra-speaker" commonalities (properties that are shared across regional and social speech communities or communities of practice). I touch upon how identifying talkers and interpreting their social characteristics might interact with recognising words and linguistic structures, and offer brief speculations about the neural processing streams that might support these inter-related tasks.

[1] Norris, D., McQueen, J.M., & Cutler, A. (2003). Perceptual learning in speech. Cog. Psych.,47, 204–238.
[2] Laver, J. (1980). The Phonetic Description of Voice Quality. Cambridge University Press.

[3] Kraljic, T., & Samuel, A.G. (2006). Generalization in perceptual learning for speech. Psychon. Bull. & Rev.13, 262–268.
[4] Smith, R. & Hawkins, S. (2012). Production and perception of speaker-specific phonetic detail at word boundaries. J. Phonetics, 40, 213-233.
[5] Barden, K.J. (2011). Perceptual learning of context-sensitive phonetic detail. Unpublished PhD thesis, Univ. Cambridge.
[6] Barden & Hawkins (under review). Perceptual learning of phonetic information that indicates morphological structure.
[7] Hawkins, S., & Smith, R.H. (2001). Polysp: A polysystemic, phonetically-rich approach to speech understanding. Italian Journal of Linguistics-Rivista di Linguistica, 13, 99–188.
[8] Creel, S.C., & Tumlin, M.A. (2011). On-line acoustic and semantic interpretation of talker information.
J. Mem. Lang., 65, 264-285.
[9] Kraljic, T., Samuel, A.G., & Brennan, S.E. (2008). First impressions and last resorts: How listeners adjust to speaker variability. Psychological Science, 19, 332-338.

Cognitive and psychological influences on phonetic adaptation in dialog

Natalie Lewandowski and Antje Schweitzer

Institut für Maschinelle Sprachverarbeitung (IMS), Universität Stuttgart
natalie@ims.uni-stuttgart.de, antje.schweitzer@ims.uni-stuttgart.de

Phonetic adaptation in communicative interaction involves either an increase (convergence) or a decrease (divergence) of similarity in the pronunciation of two dialog partners. So far, the literature on accommodation in dialog proposed three explanations for these adaptation phenomena: biologically funded accounts with a rather automatic "alignment" mechanism (Pickering & Garrod 2004), socially-motivated accounts incorporating the evaluation of the speaking partner (e.g. Giles & Ogay 2006), and hybrid accounts, viewing adaptation as biologically grounded but nevertheless susceptible to many influences and subject to considerable individual variation (Krauss & Pardo 2004, Lewandowski 2012).

Studies on phonetic convergence oftentimes report only very subtle effects or results with striking variation amongst participants. The evaluation of the conversational partner, in terms of liking, attractiveness, stereotypes (e.g. Abrego-Collier et al. 2011, Babel 2009), and also of the situational context (e.g. professional vs. private context, talker role and status) could be confirmed as some of the impacting factors. However, there still remains a lot of individual variation in the speakers' behavior, which is not accounted for, and most likely cannot be attributed to solely external factors.

A recent study on phonetic convergence in a native-nonnative context (German native speakers talking English with two English native speakers) confirmed the significant influence of the factor "phonetic talent" on the participants' degree of pronunciation adaptation (Lewandowski 2012). The analyses were based on amplitude envelope measurements at word-level during the dialogs. Although talent was a crucial factor for convergence in the dialogs, the in-group variation amongst the subjects was still considerable. A post-hoc analysis of available psychological data available for the German participants revealed further factors: the feature "openness", and to a lesser extent also "agreeableness" correlated positively with the degree of convergence of the talkers. In addition to that, the degree of convergence of the British native speaker was also mediated by her partners' personality scores, a.o. on extraversion (sign. negative correlation) and neuroticism (sign. positive correlation), while these features were not influencing the second English native speaker in his behavior towards the same subjects. This suggests that not only the speaker's own talent and personality can influence phonetic convergence but possibly also the personality estimation performed of the current conversational partner, reaching beyond simple social liking judgments.

Inter-individual variability in speech adaptation to noisy environments

Maëva Garnier[1], Nathalie Henrich[1], Lucie Ménard[2], Gabrielle Richard[2]

[1] GIPSA-lab, Grenoble, France
[2] UQAM, Montreal, Canada

Speakers modify their speech production when communicating in noisy environments. This speech adaptation, commonly referred to as the Lombard effect, mainly consists in the increase of vocal intensity (Lombard, 1911). Many other speech modifications are observed though, which can be interpreted as listener-oriented strategies aiming at compensating for the degraded speech intelligibility (Cooke et al., 2013; Garnier et al., 2010; Junqua, 1993). A first group of modifications observed may aim at improving speech detection and segregation from the background noise (increased intensity, reallocation of speech energy as a function of the noise spectrum, enhanced speech modulations in frequency and amplitude) (Garnier and Henrich, submitted; Lu and Cooke, 2009). A second set of modifications may aim at improving the clarity of phonetic cues and at facilitating their decoding by the listener (vowel hyperarticulation, decreased speech rate, enhancement of prosodic cues to utterance parsing, …) (Garnier, 2008; Garnier et al., 2006; Garnier et al., 2012). Finally, a last set of observed modifications may aim at facilitating information retrieval by emphasizing the most important parts of the utterance (emphasis on content words)(Garnier, 2007).

In this communication, we will show how speakers do not use all these different strategies in a similar way, when coping with noisy situations of communication. From the data of several experiments on Lombard speech, involving semi-spontaneous speech interactions, we will first show how speakers differ 'quantitatively', by demonstrating various degrees of speech modification when they adapt from quiet to noisy situations. We will then show how they also differ 'qualitatively', using different combinations of the above mentioned strategies to make themselves understood in noise. From our data, two main behaviors were observed : a 'shouting' behavior that primarily aims at improving speech audibility, and a 'clarity' behavior that privilege strategies of segmental and prosodic hyperarticulation. In particular, we will show how speakers differ in their use of the visual modality to cope with the degraded audio channel, some speakers enhancing visible articulatory movements, whereas others do not exploit the visible modality that much.

We will then discuss this inter-individual variability in speech adaptation in terms of communication efficiency and production efforts. A better understanding of these aspects could help preventing or rehabilitating vocal troubles in professionals who use their voice all day in noisy environments (teachers, waiters, etc…). We also expect discussions with the participants of the summerschool on the possible factors of this interspeaker variability.

References:

Cooke, M., King, S., Garnier, M., and Aubanel, V. (2013). The listening talker: a review of human and algorithmic context-induced modifications of speech. Computer Speech and Language.

Garnier, M. (2007). Communiquer en environnement bruyant : de l'adaptation jusqu'au forcage vocal. (Paris, University Paris 6).
Garnier, M. (2008). May speech modifications in noise contribute to enhance audio-visible cues to segment perception? In AVSP (Moreton Island, Australia).
Garnier, M., Dohen, M., Lœvenbruck, H., Welby, P., and Bailly, L. (2006). The Lombard Effect: a physiological reflex or a controlled intelligibility enhancement? In 7th International Seminar on Speech Production (Ubatuba, Brazil), pp. 255–262.
Garnier, M., and Henrich, N. (submitted). Speaking in noise: does the Lombard effect improve acoustic contrasts between speech and ambient noise? Computer Speech and Language.
Garnier, M., Henrich, N., and Dubois, D. (2010). Influence of Sound Immersion and Communicative Interaction on the Lombard Effect. Journal of Speech, Language and Hearing Research 53, 588-608.
Garnier, M., Ménard, L., and Richard, G. (2012). Effect of being seen on the production of visible speech cues. A pilot study on Lombard speech. In Interspeech 2012 (Portland, OR).
Junqua, J.C. (1993). The lombard reflex and it role on human listener and automatic speech recognizers. Journal of the Acoustic Society of America 93, 510-524.
Lombard, E. (1911). Le signe de l'elevation de la voix. Annales des maladies de l'oreille et du larynx 37, 101-119.
Lu, Y., and Cooke, M. (2009). Speech production modifications produced in the presence of low-pass and high-pass filtered noise. The Journal of the Acoustical Society of America 126, 1495-1499.

Effect of word position on stop bursts in Pitjantjatjara.

Marija Tabain[1] and Andy Butcher[2]

[1]La Trobe University, Melbourne, Australia
[2]Flinders University, Adelaide, Australia

Recent work on the lexical prosody of Pitjantjatjara has shown that the initial syllable of the word is the location of lexical stress, with no evidence for secondary stress elsewhere in the word (Tabain & Fletcher 2012). The initial syllable in Pitjantjatjara is marked by greater duration, a word boundary tone aligned with the left edge, and possibly by greater RMS energy. Notably, there is little evidence for vowel reduction, and no evidence for effects of spectral tilt on the vowel.

In the present study, we seek to determine whether there are effects of word-initial position on the stop burst in Pitjantjatjara, which has five places of articulation in the stop series /p t ʈ c k/. This work is motivated by what Butcher (2006) has termed the "place of articulation imperative" in Australian languages – namely, consonants exhibit various behaviours, such as blocking of coarticulation or carrying of prosodic cues, if this is beneficial to maintaining place of articulation contrasts. We analysed 8605 oral stop bursts (3676 word-initial vs. 4929 word-medial) produced by 9 different speakers. The data come both from read texts (two different stories), and from readings of phonologically balanced word lists. We selected the spectral range from 1-6 kHz, and used nine different measures to describe the spectral data: F2, F3, F4, spectral tilt, 1st spectral moment (mean, or centre of gravity – COG), 2nd spectral moment, 3rd spectral moment (skewness), and 4th spectral moment (kurtosis). In addition we measured burst duration, giving 10 measures in total. Data were analysed separately for each of the three following vowel contexts /i a u/. We conducted Linear Mixed Effects analyses of the data, with speaker and wordlist/text set as random factors.

Of the 15 tests conducted for each spectral or duration measure (5 stops * 3 vowels), spectral tilt and the 1st and 3rd spectral moments (COG and skewness) had the largest number of effects significant at the $p < 0.01$ level – there were eight or nine significant effects for these measures, compared to between three and five significant effects for the formant measures and the other two spectral moments. The significant differences in burst duration are to be expected as strengthening effects of word-initial position, with greater burst duration in this position – this was true for /p c k/. However, this was not true for the apicals /t ʈ/, which neutralize in initial position to what may be labelled /T/ (the only significant differences here were in the context of /i/, but in this case the initial neutralized /T/ was in fact a few milliseconds shorter than the medial /t/ and /ʈ/).

As regards the spectral measures tilt, COG and skewness, statistically significant effects were more likely in the context of /a/ (13 significant effects vs. eight for /i/ and only five for /u/ - in this last case, all of the effects were for /c/ and /k/). Significant effects were also more likely for /k/ (eight significant effects vs. four or five for the other four places of articulation). For /p/ in the context of /a/ (and to a lesser extent /i/), tilt and COG are higher in word-initial position, while skewness is lower – this reflects a slight increase in

energy in higher frequency parts of the spectrum relative to lower frequency parts of the spectrum. Similar results apply for /t/ and /ʈ/ in the context of /a/ and /i/. The results for /c/ are somewhat different to /p t ʈ/, in that significant results occurred for /u/, but not for /a/ or /i/ (with one exception, a lower skewness value for /c/ in the context of /a/, as is also the case for /p t ʈ/ - a lower skewness value means that the spectrum is less right-skewed). For /c/ in the context of /u/, the reverse pattern emerges: tilt and COG are *lower* in word-initial position, while skewness is *higher*. These results suggest that in this case the balance of spectral energy has moved to lower frequencies in the spectrum. When we finally turn to /k/, which has the largest number of significant results, the different patterns in the context of /u/ vs. /a/ and /i/ re-emerge: in the context of /a/ and /i/, spectral tilt and COG are lower in word-initial position for /k/, while skewness is higher; but in the context of /u/ the reverse is true, with tilt and COG being higher (there is no significant effect on skewness). These results for /k/ are different to all the other consonants, in that the balance of spectral energy moves to lower frequencies in the spectrum in word-initial position in the context of following /a/ and /i/, whereas the opposite appears to be the case for /u/.

It is not quite clear why these differences arise in spectral tilt, COG and skewness. The direction of the effect (i.e. whether the balance of spectral energy is in lower or higher parts of the spectrum) is dependent on consonant place of articulation and on following vowel context, with the back consonant /k/ behaving differently from the other consonants, and the back vowel /u/ behaving differently from the other vowel contexts. Since the spectral measures behave differently when there is a constriction towards the back of the oral cavity, it is possible that there is a particular interaction between build-up of intraoral pressure and excitation of the relevant cavity at the moment of stop release; it is further possible that this interaction is dependent on both supralaryngeal adjustments in word-initial position, and on glottal source characteristics in word-initial position. It is also not clear if these effects of word-initial position serve to enhance the contrast between different stop places of articulation, or if they simply serve to mark the word-initial syllable in Pitjantjatjara in the same way that spectral tilt marks the vowel in stressed syllables in some Germanic languages. Much further work is needed to clarify the source of these results. It is, however, clear that there are effects of word-initial vs. word-medial position on the spectral burst of the stop in Pitjantjatjara.

References:

Tabain, M. & Fletcher, J. (2012). A preliminary study of intonation in Pitjantjatjara. *14th Australasian International Conference on Speech Science and Technology*. Sydney: Australia, 213-216.

Butcher, A.R. (2006). "Australian Aboriginal Languages: Consonant- Salient Phonologies and the 'Place-of-Articulation Imperative'." In *Speech Production: Models, Phonetic Processes, and Techniques*. New York, USA: Psychology Press, pp. 187-210.

Speech perception and listener adjustments

Frank Eisner

MPI for Psycholinguistics Nijmegen

To understand speech, listeners need to adjust to a variety of factors that exert influence on its acoustic-phonetic properties. These factors include differences between talkers (e.g. vocal tract anatomy, regional accent), differences within talkers (e.g. register, physiological state), and general signal quality (e.g. background noise, filtering). The impact they have on the speech signal can be detrimental to the performance of automatic speech recognition systems, yet human listeners are usually able to adjust quite easily. An emerging body of research aims to understand the cognitive mechanisms that underlie this adaptability of the perceptual system. This work has revealed learning processes, which can act fast and induce long-lasting changes in the mapping of acoustical cues in the speech signal onto linguistically meaningful units.

Because perceptual learning can become effective quickly, it is relatively easy to study in a laboratory setting. Perceptual adjustments to various sources of variability have been observed after short exposure periods on the order of minutes or hours. The dependent measure in such experiments is typically a perceptual shift, for instance to a phoneme category boundary, or a global increase in intelligibility, for instance measured at the lexical level. In this talk I will discuss some recent studies that have used these learning paradigms in order to understand basic properties of the adaptation process – when it occurs, what constrains it, how general or specific it is, and what kinds of information in the speech signal can drive it.

Variability is naturally present in speech and the ability of the perceptual system to adjust to it is essential for speech comprehension. Understanding the mechanisms, which enable this adaptability thus gives us a more complete picture of spoken-language processing. I will discuss some implications of this recent literature on cognitive models on speech perception, and outline some potential areas in which knowledge about perceptual learning processes can guide practical applications.

References:

Samuel, A. G., & Kraljic, T. (2009). Perceptual learning for speech. Attention, Perception & Psychophysics, 71(6), 1207–1218. doi:10.3758/APP.71.6.1207

Cutler, A., Eisner, F., McQueen, J. M., & Norris, D. (2010). How abstract phonemic categories are necessary for coping with speaker-related variation. Labphon 10, 10, 91–111.

Reinisch, E., Weber, A., & Mitterer, H. (2012). Listeners Retune Phoneme Categories Across Languages. Journal of Experimental Psychology: Human Perception and Performance, 39(1), 75–86. doi:10.1037/a0027979

Listener-specific perception of speaker-specific productions? Evidence from intonation and supralaryngeal articulation across focus structures in German

Francesco Cangemi

Universität zu Köln, Université de Toulouse II Le Mirail
fcangemi@uni-koeln.de

A large number of recent *production* studies shows that speakers of a same variety might use different strategies to encode (phonetically) a given (phonological) contrast. For example, it appears that the contrast between surprise and assertiveness in German is signalled by f0 contours with different peak alignment for some speakers and different global shapes for other speakers.

It is not an overstatement to say that phonetic research in recent years has stepped away from brutal averaging of data elicited from different speakers, as the use of increasingly more refined statistical modelling shows as well. Some areas in phonetics have been characterized from their very beginning by a sheer attention to speaker-specific strategies, as in the case of articulatory phonetics, perhaps due to the usually relatively small number of subjects participating to data collection. However, other areas which have traditionally relied on more abundant data are also starting to take speaker-dependent variability into account, as in the case of intonation research cited above.

Notable exceptions exist, but research on individual systematic variations in *perception,* on the other hand, is comparatively less copious. In particular, studies on the effects of speaker-specific production strategies on perception are rare. That is, a given phonological contrast might well be phonetically implemented through different speaker-specific production strategies, but this does not necessarily mean that listeners will perceive equally well the contrasts expressed through different strategies. In other words, some production strategies might be more perceptually robust than others – some speakers might well be more clearly understandable than others. By reversing this perspective, however, we might also expect some listeners to be more sensitive than others to a given speaker-specific production strategy. That is, if we take into account speaker-specific production strategies, we must also allow for listener-specific perception strategies, and for interactions between the two.

In this contribution, we adopt this perspective in examining previously collected data concerning the effects of focus type on supralaryngeal articulation and intonation in German. In a production study, five speakers read sentences composed by the same lexical material, but in which the target word appeared in four different focus conditions. Lip aperture trajectories were measured and pitch accent was labelled on the stressed syllable of the target word. Results show that different focus types are produced with different lip aperture trajectories only by some of the speakers. Pitch accent choice across focus types is also variable across speakers, with one speaker using only one pitch accent type for all focus types, and two speakers using three pitch accent types. In a subsequent perception study, twenty listeners categorized focus type of the originally

produced stimuli. Results show that listeners are generally able to retrieve the intended focus structure, but with a two-fold variation: some listeners perform better than others (20% variation in correct categorisation), and some speakers produce items which are more easily categorised (10% variation in correct categorisation).

The focus of this contribution will be on the exploration of the interplay between speakers' productions and listeners' categorisation judgments, and specifically on the possible correlations between the predominant use of intonational or supralaryngeal cues in both production and perception, thus linking speaker-specific and listener-specific strategies. This is currently being tested through a new perception study which involves f0-resynthesis of the original stimuli, and which would greatly benefit from discussion with the summer school's invited speakers and participants on statistical modelling *inter alia*.

Individual-specific prosodic encoding and decoding of informativity

Iris Chuoying Ouyang and Elsi Kaiser

Department of Linguistics at the University of Southern California, USA

Across languages, it is widely accepted that prosodic structure can convey the extent to which a linguistic unit is informative (e.g. word frequency: [1]; contextual probability: [2]; focus: [3]). However, relatively little attention has been paid to the fact that language users vary in the prosodic cues they tend to produce and perceive (based on data from our previous work: [4][5]). To investigate the role that inter-individual variability plays in the processing of prosody with respect to informativity, we are conducting a study that consists of two experiments: (i) a production experiment that assesses between-speaker variability in the prosodic cues for word frequency, contextual probability, and focus, and (ii) a perception experiment that examines whether an individual's speaking style matches their own perceptual preference. The acoustic measures in this study include duration, pitch, and formant values.

The production experiment uses an interactive design, where participants (N>=16) work in pairs, asking and answering each other questions in turn (48 target trials, 48 filler trials). Three types of questions are used to elicit different **foci** in the answers: VP focus (e.g., A: *What did Anna and Mary do in the sea?* B: *They found fish.*), new-information narrow focus (e.g., A: *What did Anna and Mary found in the sea?* B: *They found fish.*), and corrective narrow focus (e.g., A: *I've heard Anna and Mary found boats in the sea.* B: *No, they found fish.*). Target nouns differ in **word frequency** (high, e.g. *fish*/*trees*, or low, e.g. *shells*/*lions*) and **contextual probability** (high, e.g. *fish*/*shells*, or low, e.g. *trees*/*lions*). Based on existing findings, we expect that longer duration, larger pitch excursion, and more extreme formant values appear when a target noun is narrowly focused, lexically infrequent, and contextually improbable. One question worth asking is: Are there speaker-specific behaviors in terms of which acoustic dimensions a participant uses to mark different kinds of informativity? In other words, we know that speakers have their own 'favorite' acoustic dimensions – some people might never substantially vary their formant values, and some others might stick to a certain pitch ranges – but for those who do produce a particular prosodic cue (e.g. lengthening), do they produce it for all three kinds of informativity that are examined in this study (i.e. narrow focus, low word frequency, and low contextual probability)? Could it be the case that some speakers consistently lengthen lexically infrequent words but expand pitch ranges for contextually improbable words, while others vary pitch ranges to indicate word frequency but change durations to mark contextual probability?

The perception experiment will use recordings of the answers (e.g. *they found fish.*) collected in the production experiment and have the same participants come back and listen to sentences produced by unfamiliar voices (i.e. they will not listen to sentences produced by themselves or their partner in the production experiment). Participants will be asked to guess which of the three types of questions (i.e. VP focus, new-information narrow focus, and corrective narrow focus) that a sentence was meant to respond to. We expect that participants' performance levels in this identification task will differ between

individuals (i.e. listeners) depending on who spoke the sentences, which will reveal what kinds of prosodic cues a listener attends to. By cross-comparing the results of these two experiments, We will be able to see if the participants' own production tendencies are in any kind of relationship with their perceptual biases, for example, do those who do not lengthen focused words also ignore the duration cue produced by others, or the opposite?

One of the challenges in the research of language processing is to explain the amazing ability that the language production and comprehension systems have to communicate with completely unfamiliar voices without an apparent 'adaptation phase'. The findings of this study will shed light on the mechanisms involved in this process.

References:

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

[2] Pan, S. and Hirschberg, J. (2000) Modeling Local Context for Pitch Accent Prediction. In *Proceedings of the Annual Meeting of Association for Computational Linguistics, Hong Kong, 2000*.

[3] Breen, M, Fedorenko, E., Wagner, M., and Gibson, E. (2010). Acoustic correlates of information structure. *Language and Cognitive Processes*, 25: 7, 1044-1098.

[4] Ouyang, I. C. and Kaiser, E. (2013). Prosody and information structure in a tone language: an investigation of Mandarin Chinese. *Language and Cognitive Processes*, DOI:10.1080/01690965.2013.805795.

[5] Ouyang, I. C. (In press). Non-segmental cues for syllable perception: the role of local tonal f0 and global speech rate in syllabification. *International Journal of Computational Linguistics and Chinese Language Processing*.

Psycholinguistics and Planning

Benjamin Swets

Grand Valley State University, Allendale

Traditional psycholinguistic theories of language production have held that speakers plan linguistic material in an inflexibly incremental manner. On this view, all speakers will plan the same small amount of information (e.g., a clause or a phonological phrase) prior to articulation, regardless of the contextual circumstances. An emerging view entertains the possibility that planning sentences is a more flexible and adaptive affair (see all references below). For example, speakers addressing actual people in a dialogue will take an equal amount of time to plan more content than speakers without addressees, thus increasing their planning efficiency (Swets, Jacovina & Gerrig, 2013). Speakers under time pressure will plan less content prior to articulation than speakers without any time pressure, thus modulating their scope of planning (Ferreira & Swets, 2002; 2005). Such results reveal the manner in which external, situational contexts elicit differences in planning scope. A relatively new domain of research is to investigate how organism-internal, cognitive contexts that are specific to individual speakers might interact with or constrain the cognitive processes that guide speech production.

I will present research examining individual differences in speech planning. In doing so, I will review literature on the flexibility of planning scope in language production as well as individual differences in language processing more generally. Then I will present some of the research that connects planning variability to working memory variability (Swets, Jacovina & Gerrig, in press). In this research, speakers described picture arrays to partners in a matching game. The arrays sometimes required speakers to note a contrast between a sentence-initial object (e.g., a four-legged cat) and a sentence-final object (e.g., a three-legged cat). We measured speakers' working memory prior to the session, then we recorded their eye movements during speech. The eye-movement measures revealed that speakers with higher working memory capacity were more likely to gather a larger amount of visual information prior to speech. They did so by being more likely to look ahead and fixate the object to be described sentence-finally. These speakers were also more likely to reference the contrast early in speech (e.g., by telling their partners to move "the cat with three legs" rather than simply "the cat"), reflecting an increased capacity both to gather and to encode material linguistically prior to speech.

Such analyses of individual differences can help to develop theories about the nature of the cognitive system that processes speech. For example, they afford theoretical developments regarding the mechanistic role that working memory plays in speech planning, as well as the role that working memory might play in language processing more generally. I will also suggest some future directions in which more research might help expand our understanding of language processing by examining individual differences, and also explore some limitations of the individual differences approach.

References:

Ferreira, F. & Swets, B. (2002). How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language*, 46, 57-84.

Ferreira, F., & Swets, B. (2005). The production and comprehension of resumptive pronouns in relative clause "island" contexts. In A. Cutler (Ed.), *Twenty-first Century Psycholinguistics: Four Cornerstones* (pp. 263-278). Mahway, NJ: Lawrence Erlbaum Associates.

Fuchs, S., Petrone, C., Krivokapic, J., & Hoole, P. (2013). Acoustic and respiratory evidence for utterance planning in German. *Journal of Phonetics*, 41, 29-47.

Swets, B., Jacovina, M. E., & Gerrig, R. J. (2013). Effects of conversational pressures on speech planning. *Discourse Processes*, 50, 23-51.

Swets, B., Jacovina, M. E., & Gerrig, R. J. (in press). Individual differences in the scope of speech planning: Evidence from eye movements. Language and Cognition.

Wagner, V., Jescheniak, J. D., & Schriefers, H. (2010). On the flexibility of grammatical advance planning during sentence production: Effects of cognitive load on multiple lexical access. Journal of Experimental Psychology: Learning, Memory, and Cognition, 36, 423-440.

Articulatory Motor control in adults who stutter – An electromagnetic articulatory study

Cornelia Heyde

Queen Margaret University, Edinburgh

In my project I will study the articulation of people who stutter aiming to explore the nature of disfluencies. In contrast to prior studies which primarily focus on the differences in voicing/phonation onset between people who stutter (PWS) and people who do not stutter (PNS), the present study will focus on articulatory onset. The relation of articulatory reaction time (ART) and voice reaction time (VRT) will be looked at and compared between and within the groups of PWS and PNS. The use of electromagnetic articulography (EMA) will allow imaging the exact movements of the tongue and the lips and comparing results to the participants' onset of voicing.

Similar to findings of prior studies, differences in VRT between PWS and PNS are expected. When comparing ART between PWS and PNS, however, results may reveal no such difference. The delay in phonatory onset (VRT) in PWS is suggested to result from difficulty in initiation of phonation when passing from the consonant to the adjacent vowel, rather than struggling to initiate articulation of the initial consonant (ART). These difficulties in getting away from the initial consonant are expected to be found in non-fluent as well as fluent PWS speech. In addition to difficulties in timing, the issue of coordination of the articulators will be investigated regarding the participants' motor coordination in fluent as well as disfluent speech.

The purpose of the study is expressed in a quote by Paul Brocklehurst (2013) who states: "*The elephant in the room here is the fact that we do not yet know for sure the mechanism that gives rise to stuttered disfluencies. This continued uncertainty leads us to an unsatisfactory overreliance on perceptual factors as a basis for definition*".

Research questions will address:

- Do results suggest difficulties in initiation of articulation in PWS, i.e. longer ART in PWS compared to PNS?
- Do PWS struggle to initiate phonation, i.e. longer VRT in relation to ART in PWS than PNS?
- Can evidence be found for deficits in lingual motor coordination, e.g. longer durations and more variable distance travelled by the tongue for PNS than PWS?
- Do PWS present with difficulty initiating coarticulation in disfluent speech, i.e. no sign of coarticulation immediately prior to blocks?

Feedback methods to improve phonetic and phonological skills in foreign language acquisition

Jeanin Jügler

Department of Computational Linguistics and Phonetics, Saarland University, Saarbrücken, Germany

Pronunciation, the articulation of sounds as well as the realization of prosody, especially the use of intonation, has been left out to a great extent in second language teaching (Hirschfeld and Trouvain, 2007:172). Though, within the last decade the attitude towards pronunciation seems to have changed step by step. The development of computer-based language learning (CALL) and especially computer-based pronunciation training (CAPT) systems obtained a special interest over the years (e.g. Euronounce (Demenko et al., 2009); Dutch-CAPT (Neri et al., 2008)). Many of these systems include an automated speech recognizer that aligns the input of a learner with the pre-recorded speech of a native speaker of a given language. On the basis of this comparison, feedback regarding any specific phonetic or prosodic error can be given. It has been proposed by researchers that feedback helps to improve learning the pronunciation of a foreign language (e.g. Precoda et al., 2000) and argued that the lack of good feedback in CAPT systems constitutes the negative facets of current implementations (Engwall et al., 2004:1693).

Therefore, the aim of this work is 1) to deal with the general concept of feedback, to describe effects of different types of feedback, not only in relation to pronunciation learning but the complete learning process, 2) to analyze and evaluate different feedback methods which have already been used in existing CALL/CAPT systems, and 3) to discuss further methods to improve phonetic and phonological skills in the foreign language. These methods will be tested and might later be implemented in a new CAPT system currently developed within a project called "Individualized Feedback in Computer-Assisted Spoken Language Learning" by researchers of Saarland University and LORIA, France. This system takes into account the German-French language pair, keeping in mind speaker-specific behavior. That is, that the included automated speech recognizer needs to deal with the factor of speaker individuality and non-native speech input. Further on, people show different learning behavior in respect to foreign language learning. Therefore, a system has to be developed which comprises a high standard of individualization, not only regarding task types and interface usage preferences but also the way feedback will be applied. One possibility of feedback integration presents the inclusion of an animated vocal tract in order to provide a better articulatory illustration of difficult sounds. This might constitute a beneficial effect in comparison the mere presentation of an oscillogram contrasting the learner's and native speaker's output, plain text explanation or a numbered scale which might be very difficult for the learner to interpret and to execute phonetic or prosodic alterations.

References:

G. Demenko, A. Wagner, N. Cylwik, and O. Jokisch. An audiovisual feedback system for acquiring L2 pronunciation and L2 prosody. In *2nd ISCA Workshop on Speech and Language Technology in Education (SLaTE)*, pages 2–5, 2009.

O. Engwall, P. Wik, J. Beskow, and B. Granström. Design strategies for a virtual language tutor. In S. H. Kim and D. H. Young, editors, *Proc. ICSLP 2004*, pages 1693–1696, 2004.
U. Hirschfeld and J. Trouvain. Teaching prosody in German as a foreign language. In J. Trouvain and U. Gut, editors, *Non-Native Prosody. Phonetic Description and Teaching Practice. (Trends in Linguistics. Studies and Monographs (TiLSM) 186)*, pages 171–187, 2007.
A. Neri, C. Cucchiarini, and H. Strik. The effectiveness of computer-based corrective feedback for improving segmental quality in L2-Dutch. In *ReCALL*, volume 20, pages 225–243, 2008.
K. Precoda, Chr. Halverson, and H. Franco. Effects of speech recognition-based pronunciation feedback on second-language pronunciation ability. In *Proc. Integrating Speech Technology in (Language) Learning (InSTILL)*, pages 102–105, 2000.

Articulatory normalization via imitation strategy in phone classification task

Claudia Canevari

Istituto Italiano di Tecnologia – RBCS Genova – Italy
claudia.canevari@iit.it

Automatic Speech Recognition (ASR) systems classically apply normalization strategies that minimize inter-speaker variability by explicitly removing speakers' peculiarities (e.g. through Cepstral mean removal, Anastasakos et al., 1994, or vocal tract length normalization, Eide and Gish, 1996) or adapting different speakers to a reference model (e.g. Huang, 1992) or creating compact adaptable and robust models (Anastasakos et al., 1996) without exploiting speakers' variations in learning and recognition processes.

Here we propose a speaker normalization strategy that uses measured articulatory information and test it in a phone classification task. Such approach is in line with a new trend in Automatic Speech Recognition where speech production knowledge is combined with the traditional acoustic features. When measured articulatory data are used, those data are typically only available during training, while during recognition they can be recovered from speech acoustics through an inverse procedure, the so called Acoustic-to-Articulatory Mapping (AAM).

We first learnt a speaker-independent AAM where the speech acoustics of several training (seen) speakers were mapped onto the motor plans of one single subject, thus simulating an infant trying to mimic utterances produced by other speakers. We then evaluated the utility of the infant's reconstructed articulatory information, when classifying the speech of speakers whose data were either used or not used for training.

Preliminary results suggest that our articulatory normalization produces a more accurate phone classification (i) w.r.t. the acoustic baseline, where no articulatory information have been used and no normalization strategy have been applied and (i) w.r.t. a "corresponding" acoustic normalization strategy where the normalization is carried out on the acoustic domain only. Further studies will be necessary to assess whether and how articulatory similarity between infant and speakers can affect the phone discrimination.

The impact of source-filter interaction on speaker's formants and pitch variability

Nathalie Henrich and Maëva Garnier

GIPSA-lab, Grenoble, France

The variability of formants and pitch are two key aspects of speech production. In the classical framework of the source-filter theory (Fant, 1960), formant frequencies vary as a function of articulatory modifications, and pitch depends mainly on vocal-fold biomechanical configuration. In this paper, we aim to detail the different aspects of source-filter interaction in human speech production, and to discuss how it could induce speaker-specific differences.

In phonetic sciences and speech processing, the contribution of vocal tract to the radiated speech sound is assumed to be linear and independant of glottal source properties. Yet, this source-filter theory is a first-order approximation of the physics of speech production, as glottal source and vocal-tract transfert function are physically coupled (Fant and Lin, 1987; Klatt and Klatt, 1990; Childers and Wong, 1994; Titze and Story, 1997; Titze, 2004a, 2004b, 2008a; Titze et al. 2008b). In several ways, the vocal-tract transfer function is dependent on glottal source properties. Conversely, pitch can be slightly modified by vocal-tract configurations. On the basis of in vivo and in vitro experiments, we will illustrate the dependency of first-formant frequency F1 on glottal boundary conditions, vocal effort and voice quality (Barney et al., 2007 ; Garnier et al., 2008, Swerdling et al., 2010). The impact of vocal-tract constrictions on pitch will be demonstrated (Bailly et al., 2008). We will discuss the need to take into account source-filter interaction phenomena in the interpretation of formant and pitch variations within and between speakers.

References:

Bailly L., Pelorson X., Henrich N., Ruty N. (2008) Influence of a constriction in the near field of the vocal folds: Physical modeling and experimental validation, J. Acoust. Soc. Amer., Vol. 124 (5), pp. 3296-3308.

Barney A., De Stefano A. and Henrich N. (2007) The effect of glottal opening on the acoustic response of the vocal tract, Acta Acustica united with Acustica, vol. 93 (6), pp. 1046-1056.

Childers, D. G. & Wong, C. F. (1994) Measuring and modeling vocal source-tract interaction., IEEE Trans Biomed Eng, vol. 41, n°7, pp. 663-671.

Fant G. (1960) Acoustic theory of speech production. Mouton, La Hague

Fant, G. and Lin, Q. (1987) Glottal source - vocal tract acoustic interaction, STL-QPSR, vol. 28, n°1, pp. 13-27.

Garnier, M., Wolfe, J., Henrich, N., Smith, J. (2008) Interrelationship between vocal effort and vocal tract acoustics: a pilot study. In Proceedings ICSLP, Brisbane.

Klatt, D. & Klatt, L. (1990) Analysis, synthesis, and perception of voice quality variations among female and male talkers, J. Acous. Soc. Am., vol. 87, n°2, pp. 820-857.

Titze I.R. and Story B.H. (1997) Acoustic interactions of the voice source with the lower vocal tract. J. Acoust. Soc. Am., 101 (4), pp. 2234-2243.

Titze, I. R. (2004a) A theoretical study of F0-F1 interaction with application to resonant speaking and singing voice., J Voice, vol. 18, n°3, pp. 292-298.

Titze, I. R. (2004b) Theory of Glottal Airflow and Source-Filter Interaction in Speaking and Singing, Acta Acustica united with Acustica, vol. 90, pp. 641-648.

Titze, I. R. (2008a) Nonlinear source-filter coupling in phonation: theory. J Acoust Soc Am, vol. 123, pp. 2733-2749

Swerdling Y., Smith J., Wolfe J. (2010) The effect of whisper and creak vocal mechanisms on vocal tract resonances. J Acoust Soc Am, vol. 127(4), pp. 2590-2598.

Phonetic realisation of linking and intrusive /r/ in Australian English

Ivan Yuen, Felicity Cox, Katherine Demuth

Linguistic Department, CCD, Macquarie University

Word pairs like 'lore/law' in Australian English are homophones in citation, but they differ in connected speech, specifically before a vowel. Cox & Buckley (2009) showed that Australian speakers managed hiatus between two adjacent vowels across word-boundary by means of epentheticr. However, the word pair also differs in spelling. Assuming a letter-to-sound rule in lexical representation, this raises the question of whether orthographic representation will impact on phonetic realisations of the word pair. There are two scenarios: (1) do they have the same phonetic realisation, or (2) do they have different phonetic realisations. In addition, we would investigate whether morphological structure will interact with orthography on how epenthetic-r will be realised? To address this question, we would conduct a production experiment looking at monosyllabic word pairs (e.g. lore/law), non-morphological disyllabic word pairs (e.g. pastor/pasta), and morphological disyllabic word pairs (e.g. beater/beta) in three vowel conditioning contexts: (1) before stressed vowel; (2) before unstressed vowel and (3) utterance-finally. We would predict that epenthetic-r will manifest pre-vocalically (i.e. before stressed and unstressed vowels); however, no epenthetic-r will manifest utterance-finally.

A preliminary analysis showed no epenthetic-r in utterance-final position as predicted. Surprisingly, r-insertion was not a primary strategy in dealing with vowel hiatus, with only 12 cases out of 100 items. Instead, a majority of 86 items exhibited the use of glottalisation/glottal stops to break up two contiguous vowels, questioning the phonological status of -r.

Speaker-specificity in the time and f0 domain

Adrian Leemann[1], Volker Dellwo[1], Hansjörg Mixdorff[2], Marie-José Kolly[1]

[1]Phonetics Laboratory, Department of General Linguistics, University of Zurich
[2]Beuth-Hochschule für Technik, Berlin
adrian.leemann@pholab.uzh.ch, marie-jose.kolly@pholab.uzh.ch,
volker.dellwo@uzh.ch, mixdorff@beuth-hochschule.de

Previous research has pointed out significant variation between speakers of the same language in the time domain (e.g. Laan 1997, Wiget et al. 2010, Dellwo et al. 2012, Arvaniti 2012) as well as in the f0 domain (e.g. Künzel 1987, Nolan 2002, Laboutin et al. 2007). Until today, however, a combination of the two domains for the purposes of speaker discrimination has gained little attention.

16 subjects who speak Zurich German were recorded (256 read sentences per speaker; 55'000 syllables in total). The material was segment labeled manually. To test if we find between-speaker differences in the time domain, we applied a number of temporal measures, including %V (the percentage of vocalic material in the speech signal) and %VO (the percentage of voiced material in the speech signal) (see Dellwo et al. 2012 for an overview). To examine if the speakers differ in the f0 domain, we applied the Fujisaki intonation model (Fujisaki and Hirose 1984). All of the temporal measures and f0 measures were compared between the speakers.

Results reveal that speakers behave differently from one another in both domains. We find significant effects for the factor *speaker* in all calculated models (linear mixed effect models with *speaker* as fixed effect and *sentence* and *gender* as random effects; all $p<.0001^{**}$). Four speakers' mean scores of %V and %VO as well as of four Fujisaki f0 parameters (phrase slope, phrase duration, accent height, accent duration) were subjected to a discriminant analysis (DA). We find evidence that individual speaker regions can be delineated with two discriminant functions. Results show that speakers can be discriminated relatively well with a combination of the two domains, reaching a classification rate of 63%. If only the f0 parameters are used for the DA, we obtain a rate of 57%. The DA based on a combination of f0 and timing information reveals promising findings for forensic voice comparison. In future work, further temporal and f0 measures will be included in the DAs so as to maximize between-speaker differences.

References:

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

Dellwo, V., A. Leemann, M.-J. Kolly (2012). Speaker idiosyncratic rhythmic features in the speech signal. *Proceedings of Interspeech 2012*, 9.-13.09.2012, Portland, USA.

Fujisaki, H. and K. Hirose (1984). Analysis of voice fundamental frequency contours for declarative sentences of Japanese. *Jour. of the Acoust. Society of Japan* E 5 (4): 233–242.

Künzel, H. (1987). *Sprechererkennung: Grundzüge Forensischer Sprachverarbeitung*, Heidelberg: Kriminalistik Verlag.

Laan, G. (1997). The contribution of intonation, segmental durations, and spectral features to the perception of a spontaneous and a read speaking style. *Speech Communication* 22, 43-65.

Labutin, P., S. Koval, A. Raev (2007). Speaker identification based on the statistical analysis of f0. Proc. IAFPA 2007, The College of St Mark & St John, Plymouth, UK, 2007.

Mixdorff, H. (2000). A novel approach to the fully automatic extraction of Fujisaki model parameters. Proceedings of ICASSP 2000, vol. 3, 1281-1284. Instanbul, Turkey.
Nolan, F. (2002). Intonation in speaker identification: an experiment on pitch alignment features. *Forensic Linguistics, 9(1), 1-21.*
Wiget, L, L. White, B. Schuppler, I. Grenon, O. Rauch, S.L. Mattys (2010). How stable are acoustics of contrastive speech rhythm? *JASA* 127, 1559-1569.

Separating segmental and prosodic contributions to intelligibility

Daniel McCloy

Graduate Research Associate Department of Linguistics, University of Washington
http://students.washington.edu/drmccloy

It is well known that the intelligibility of speech can vary both across individuals within styles or tasks, and within individuals across styles or tasks. Various properties of the speech signal have been shown to correlate with such differences in intelligibility, including speech rate,[5,7,8] segmental reduction or deletion,[1] vowel space size,[1,2,4,6] pitch range,[2] and pitch accent deletion.[3] However, these dimensions are rarely (if ever) manipulated independently in natural speech. This poses a challenge to understanding the sources of individual differences in intelligibility (both across individuals and across styles), and makes it difficult to know whether any particular dimension measured *causes* speech to be more or less intelligible, or merely *indexes* some other aspect of speech that is responsible for intelligibility differences.

As an alternative to measuring fine-grained dimensions of the speech signal, this research makes a broad distinction between prosodic dimensions (pitch, intensity, and duration) on one hand, and segmental content on the other. Through careful resynthesis, a corpus of parallel sentences are created that effectively hold constant either prosody or segmental content across resynthesized "talker". Highquality stimuli are achieved by hand-correction of glottal pulse epochs and semi-automated hand segmentation of syllable durations, followed by automated dynamic time warping of durations and swapping of pitch and intensity contours.

Results from a speech-in-noise task with both unmodified and resynthesized stimuli show that talkers with low intrinsic intelligibility may have relatively "good" prosody, evidenced by improvements in intelligibility when their prosody is mapped onto other talkers' waveforms. In contrast, talkers with high intrinsic intelligibility may have relatively "bad" prosody, evidenced by lower intelligibility caused by mapping their prosody onto other talkers. A linear mixed-effects regression model (controlling for signal processing distortion and variation in sentence difficulty) supports this view: patterns of coefficients for "prosodic donor" and "segmental donor" show different rankings than the overall intelligibility scores for unmodified talkers. Comparison between these patterns and *post-hoc* acoustic analyses of the stimuli allows classification of acoustic predictors based on how well they correlate with "prosodic donor" or "segmental donor" coefficient patterns.

References:

[1] Bond, Z. S., & Moore, T. J. (1994). A note on the acoustic-phonetic characteristics of inadvertently clear speech. *Speech Communication*, 14(4), 325–337. DOI: 10.1016/0167-6393(94)90026-4.

[2] Bradlow, A. R., Torretta, G. M., & Pisoni, D. B. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication*, 20(3-4), 255–272. DOI: 10.1016/S0167-6393(96)00063-5.

[3] Clopper, C. G., & Smiljanić, R. (2011). Effects of gender and regional dialect on prosodic patterns in American English. *Journal of Phonetics*, 39(2), 237–245. DOI: 10.1016/j.wocn.2011.02.006.

[4] Hazan, V., & Markham, D. (2004). Acoustic-phonetic correlates of talker intelligibility for adults and children. *The Journal of the Acoustical Society of America*, 116(5), 3108‑3118. DOI: 10.1121/1.1806826.
[5] Mayo, C., Aubanel, V., & Cooke, M. (2012). Effect of prosodic changes on speech intelligibility. Paper presented at the 13th Annual Conference of the International Speech Communication Association. In *INTERSPEECH-2012*. DOI: http://interspeech2012.org/accepted-abstract.html?id=661
[6] Neel, A. T. (2008). Vowel space characteristics and vowel identification accuracy. *Journal of Speech, Language, and Hearing Research*, 51(3), 574‑585. DOI: 10.1044/1092-4388(2008/041).
[7] Sommers, M. S., Nygaard, L. C., & Pisoni, D. B. (1994). Stimulus variability and spoken word recognition I: Effects of variability in speaking rate and overall amplitude. *The Journal of the Acoustical Society of America*, 96(3), 1314‑1324. DOI: 10.1121/1.411453.
[8] Tolhurst, G. C. (1957). Effects of duration and articulation changes on intelligibility, word reception and listener preference. *Journal of Speech and Hearing Disorders*, 22(3), 328‑334.

Input perception and phonological accuracyin adult learner Polish

Jacopo Saturno

University of Bergamo
jacopo.saturno@unibg.it, jacopo.saturno@yahoo.it

Within the field of Second Language Acquisition, the present study aims at evaluating the perceptual prominence of input items by analysing their phonological form in learner output. The objective of the research is to identify those features which pose the greatest difficulty in perceiving and reproducing target items.

Analysis is conducted on *ab initio* adult learner varieties of Polish. The data were elicited within the VILLA project (Varieties of Initial Learners in Language Acquisition: Dimroth *et alii* to appear) by exposing 31 Italian L1 learners with no previous experience of Slavic languages to a 14-hour course of Polish. All teacher speech and learner output was recorded and transcribed, so as to correlate the development of the interlanguage with the characteristics of the input; learners also took a series of tests, of which the Sentence Imitation task is considered here. This was administered at two different stages, in order to evaluate the effects of item frequency and trace the evolution of interlanguage phonology. Accuracy scores have been calculated in terms of the number of phonemes which were omitted or modified in learner output with respect to the target items; because of the possible interaction with implicit grammatical knowledge, the scores relative to case endings have been calculated and analysed separately.

Among the variables taken into account, transparency has been noted to be a strong predictor of phonological accuracy in the output. Transparent items are generally identified correctly (Rast 2010), but reproduced with slight deviations from the target which seem to be consistent among learners, most probably reflecting the influence of the L1 (Van Hell & Tanner 2012, Park & de Jong 2008) or other known languages (Hammarberg & Hammarberg 2009). Conversely, ample inter-speaker variability is observed in the case of non-transparent items. Significant effects have been noted for the parameter of frequency in the input, frequent items being generally reproduced more accurately (Rast 2008). Word order (SVO *vs.* OVS) and item position in the sentence (Peters 1985) are also taken into account. On the basis of the above findings, the phonological structure of target items is then analysed, in order to identify those features which in terms of segment inventories and syllabic structure (Altenberg 2005) seem to facilitate or hinder the learner's task. Finally, working memory is considered as a speaker-specific determinant of phonological accuracy (Juffs & Harrington 2011).

References:

Altenberg, Evelyn P. 2005. The Judgment, Perception, and Production of Consonant Clusters in a Second Language. *IRAL - International Review of Applied Linguistics in Language Teaching* 43(1). 53–80.

Dimroth, Christine, Rebekah Rast, Marianne Starren & Marzena Wątorek. forthcoming. Methods for studying the learning of a new language under controlled input conditions: The VILLA project.

Hammarberg, B. & B. Hammarberg. 2009. Re-setting the basis of articulation in the acquisition of new languages: A third language case study. *Processes in third language acquisition*, 74–84. Edimburgh: Edinburgh University Press. Originally 2005 in Hufeisen B. & Fouser R.J. (eds.), Introductory readings in L3, 11–18. Tübingen: Stauffenburg.

Hell, Janet G. Van & Darren Tanner. 2012. Second Language Proficiency and Cross-Language Lexical

Activation. *Language Learning* 62. 148–171.
Juffs, A. & M. Harrington. 2011. Aspects of working memory in L2 learning. *Language Teaching* 44(2). 137 - 166.
Park, Hanyong & Kenneth J. de Jong. 2008. Perceptual category mapping between English and Korean prevocalic obstruents: Evidence from mapping effects in second language identification skills. *Journal of Phonetics* 36(4). 704–723.
Peters, Ann M. 1985. Language segmentation: operating principles for the analysis and perception of language. In Dan I. Slobin (ed.), *The crosslinguistic study of language acquisition*., vol. 2, 1029–1067. Hillsdale, NJ: Lawrence Erlbaum Associates.
Rast, Rebekah. 2008. *Foreign Language Input: Initial Processing*. Clevedon: Multilingual Matters.
Rast, Rebekah. 2010. The use of prior linguistic knowledge in the early stages of L3 acquisition. *Internation review of applied Linguistics*(48). 159–183.

Attitudes to Different Types and Frequencies of Errors in Non-native German

Evghenia Goltsev

University of Freiburg/Research training group GRK DFG 1624/1
"Frequency effects in language"
Evghenia.Goltsev@frequenz.uni-freiburg.de

Since one of the essential functions of language is successful communication, it is important to identify the linguistic elements in the learner language that influence the communication. My project aims to investigate the way in which native speakers of German perceive and judge erroneous utterances produced by advanced and intermediate learners of German with a Russian background. The focus will be on factors such as intelligibility and acceptability as well as the listeners' attitudes to the speakers themselves.

The essential research question is: How does the frequency of modification types such as phonetic/phonological or morpho-syntactic errors influence negative evaluation (Albrechtsen et al. 1980; Llurda 1995)? The interaction of these and other aspects such as voice qualities will also be considered.

The role of the variables age, gender and educational background will also be examined in the reactions to the errors presented (Hellwig-Fábián 2007). The significance of gender will be determined for both the learners of German and the native speakers.

The methodological approach of my project generally follows a socio- and psycholinguistic attitude research paradigm. The reactions to erroneous utterances will be measured using the subjective evaluation test and the matched guise technique. A fixed number of audio stimuli will be created in which each stimulus contains a certain type of error. These will be read out by the same speaker imitating various accent types and proficiency levels (Garret 2010). The raters will be asked to evaluate the audio stimuli using several semantic differentials (Zahn/Hopper 1985).

References:

Albrechtsen, D. / Henriksen, B. / Faerch, C. (1980): Native Speaker Reactions to Learners' Spoken Interlanguage. In: Language Learning. 30, 365-396.
Garrett, P. (2010): Attitudes to Language. Cambridge.
Hellwig-Fábián, I. (2007): Deutsch mit ausländischem Akzent. Eine empirische Studie zu Einstellungen junger Deutscher gegenüber Sprechern mit ostslavischer Muttersprache. Berlin.
Llurda, E. (1995): Native-Speaker Reactions to nonnative speech. A review. In: Sintagma 7. 30, 43-51.
Zahn, Ch. / Hopper, R. (1985): Measuring Language Attitudes: The Speech Evaluation Instrument. Journal in Language and Psychology, 113- 123.

V-to-V coarticulation in French/Italian bilinguals: the role of language-specific constrains on articulatory strategies

Charlotte Graux, Mariapaola D'Imperio, Thierry Legou, Jonathan Harrington

Vowel - to - vowel coarticulation is generally conditioned by universal, bio-mechanical constraints, though language-specific constraints can also have an impact (Manuel, 1990; Marchal, 2011; *inter alia*). For instance, Tabain & Perrier (2007) suggest that a language's phoneme inventory has important effects on the articulatory strategies adopted by speakers. Hence we will here explore whether vowel-to-vowel coarticulatory patterns in French and Italian might be determined by the linguistic system and not merely by biomechanical factors. In order to separate the two factors, our target population for the study are French - Italian high level bilinguals. We will specifically compare the co-articulatory patters of /i/ and /u/ in both French and Italian, because both vowels are present in the language and they also share similar acoustic targets, thoug we do not know if they share the same underlying articulatory strategies.

We also know that stronger coarticulation is expected at faster speech rate. Hence, in our study we also manipulate speech rate (both normal and fast). Prosodic boundaries can also have an effect on gestural scores. We know that bothspatial and synamic properties of articulatory movements are affected by prosodic boundaries, and that the effects are stronger for constituents that are higher in the prosodic hierarchy (Fougeron & Keating, 1997). Perrier & Tabain (2007) found an interesting though unexpected effect of prosodic boundary for concerning /u/ articulation, in that a tighter lip constriction was found for weaker rather than stronger boundaries, due probably to the need of keeping F2 low so that contrast with /y/ would be preserved. The contrast preservation demand would not be present in Italian, since its vocalic system lacks /y/.

Specifically, here we test whether /u/ in Italian is fronted to a greater extent than in French by consonantal context, since /u/ does not contrast with /y/ in Italian, while it does in French. We also expect, only for French, that there will be a greater lip-rounding adjustment for /u/ to offset the tongue-fronting induced by a faster rate. The acoustic consequences of this greater lip-rounding adjustment will be that F2 will rise due to rate in /u/ much less in French than in Italian. Finally, we expect a stronger difference in tongue retraction and/or lip rounding due to both prosodic-induced and rate-induced shortening in the Italian than the French conditions, though we expect an overall effect of the manipulation in both languages.

In order to verify the prosodic boundary effects, our corpus includes three conditions: IP (Intonation Phrase) boundary, AP (Accentual Phrase) boundary, and AP medial condition. Comparable sentences were created for both French and Italian. The target vowel was put in a CV.CV sequence where the vowel is either /u/ or /i/, within non-words such as /bubu/ and /bibu/, etc. Each sentence was repeated five times by each of the speakers, in the two languages and in two rates, for a total of 300 sentences per subject. As of today, 2 speakers have been recorded.

A sociophonetic look at rhotic variation in Dundee

Thomas Jauriberry[1,2], Rudolf Sock[1,2], Albert Hamm[1]

[1] EA 1339 LILPA, 22, rue René Descartes 67100 Strasbourg
[2] Institut de Phonétique de Strasbourg, 22, rue René Descartes 67100 Strasbourg
t.jauriberry@unistra.fr, sock@unistra.fr, hamm@unistra.fr

Cross-linguistically and language-internally, /r/ is a particularly variable phoneme which exhibits several phonetic realisations. Scottish English is typically described as a rhotic variety of English, with an articulated nonprevocalic /r/ in words like *car, card, better, fear*. The rhotic variants are usually taps [ɾ] and approximants [ɹ], and vary according to internal factors such as phonological environment (Wells, 1982 : 411), but also age and gender (Romaine, 1978 ; Wells, 1982 : 411 ; Pukli & Jauriberry, 2011 ; Jauriberry, 2012). Non-prevocalically, recent findings indicate not only that /r/ is extremely variable, but also that a process of derhoticisation might be ongoing in this accent, leading to r-loss or vocalisation in coda position (Romaine, 1978 ; Stuart-Smith, 2007 ; StuartSmith et al., 2007 ; Lawson *et al.*, 2008 ; Llamas, 2010 ; Pukli & Jauriberry, 2011 ; Jauriberry, 2012). Yet, derhoticisation seems most advanced for working-class speakers, while the middle-class (MC) is little if not affected (Stuart-Smith *et al.*, 2007 ; Lennon, 2012). In Scotland, research mainly focused on the two greatest urban centres of Glasgow (Stuart-Smith & Tweedie, 2000 ; Stuart-Smith, 2007 ; Stuart-Smith et al., 2007) and Edinburgh (Romaine, 1978 ; Lawson *et al.*, 2008), and on neighbouring towns such as Ayr (Pukli & Jauriberry, 2011, Jauriberry, 2012) and Livinston (Lawson *et al.*, 2008). Towns further away, Gretna and Eyemouth, are small and spatially marginal as they are located along the national border with England (Llamas, 2010).

This paper presents results of acoustic analyses of both prevocalic and non-prevocalic /r/ in the speech of a selection from a corpus of 161 native speakers recorded throughout Central Scotland in 2012. 10 female speakers evenly divided according to age (students who are 18 and teachers from 42 to 50), have been recorded at the Morgan Academy, a Dundee middle-class high school, reading a wordlist containing 130 words with /r/ in various phonological and phonetic environments. The analyses confirm first that /r/ is extremely variable in this variety, ranging from alveolar trills to taps, approximants, and fricatives, but also vocalic variants and zero realisation. Second, phonological and phonetic environments are key factors in the variation of /r/, both prevocalic and non-prevocalic. Third, derhoticisation and R-vocalisation are rare for female MC speakers, and most frequent for the younger speakers, who tend to produce more reduced variants.

References:

JAURIBERRY, T., SOCK, R., PUKLI, M., & HAMM, A. (2012). Rhoticité et dérhoticisation en anglais écossais d'Ayrshire, Proceedings of the JEP-TALN-RECITAL joint conference 2012, Vol. 1, pp. 89-96.

LAWSON, E., STUART-SMITH, J. & SCOBBIE, J.M. (2008). "Articulatory Insights into Language Variation and Change: Preliminary Findings from an Ultrasound Study of Derhoticization in Scottish English" In K. Gorman (ed.) U. Penn Working Papers in Linguistics 14.2: Papers from NWAV 36, pages 102-110.

LENNON, R. (2012). A real-time sociophonetic study of postvocalic /r/ in the speech of schoolchildren in Bearsden. Unpublished Honours dissertation.

LLAMAS, C. (2010). Convergence and divergence across a national border. In Llamas, D. & Watt, D. (eds.), Language and Identities. Edinburgh: Edinburgh University Press, pp. 227-236.

PUKLI, M. & JAURIBERRY, T. (2011). "Language change in action – Variation in Scottish English",

RANAM 44, pages 83-100.
ROMAINE, S. (1978). “Postvocalic /r/ in Scottish English: sound change in progress?” In Trudgill, P. (ed.), Sociolinguistic Patterns in British English, London: Arnold, pages 144-157.
STUART-SMITH, J. (2007). “A sociophonetic investigation of postvocalic /r/ in Glaswegian adolescents” Proceedings of the XVIth International Congress of Phonetic Sciences, Saarbrücken, pages 1449-1452.
STUART-SMITH, J. & TWEEDIE, F. (2000). Accent change in Glaswegian: a sociophonetic investigation. Final Report to the Leverhulme Trust (Grant no. F/179/AX):
<http://www.arts.gla.ac.uk/STELLA/Glasgow%20accent/Report.htm>
STUART-SMITH, J., TIMMINS, C. & TWEEDIE, F. (2007). “Talkin’ Jockney: Accent change in Glaswegian”, Journal of Sociolinguistics, 11, pages 221-61.
WELLS, J. C. (1982). Accents of English. 3 volumes. Cambridge: Cambridge University Press.

The influence of the place of articulation on the speaker specificity of German phonemes

Carola Schindler and Christoph Draxler

Institute of Phonetics and Speech Processing, Ludwig Maximilians University, Munich, Germany
carola.schindler@phonetik.uni-muenchen.de, draxler@phonetik.uni-muenchen.de

In a previous study we analysed the speaker discriminating potential of the spectral moments of nasals and fricatives (Mook & Draxler, 2012) and showed that these acoustic parameters are quite speaker specific and worth further investigating. As the results for the alveolar phonemes consistently showed higher speaker specificity than those for labial phonemes, we assumed that the place of articulation has a significant effect on the speaker specificity of phonemes.

Our assumption was that the alveolar ridge requires more precision in the articulation and thus allows less variation in the production of a phoneme than the lips. Hence, the variation of a speakers articulation in these phonemes would be small causing a low intraspeaker variability, which would be useful for the speaker specificity.

For the experiment we used phonemes of 49 male speakers of German from the Verbmobil corpus. After extracting the relevant phonemes and calculating the spectra in the EMU speech database system, we imported the data into the statistic software R, where we measured the spectral moments for the midpoint of the spectrum of each segment. The obtained data was evaluated statistically by an analysis of variance.

The diagram shows that the F-ratios of the four spectral moments were always higher for the alveolar than for the labial phonemes. Further analysis contradicted our initial assumption that the intra-speaker variability would be significantly smaller for the alveolar than for the labial phonemes. Instead, inter-speaker variability was constantly higher for the alveolar phonemes. Maybe the properties of the alveolar ridge differ more between speakers than the properties of their lips.

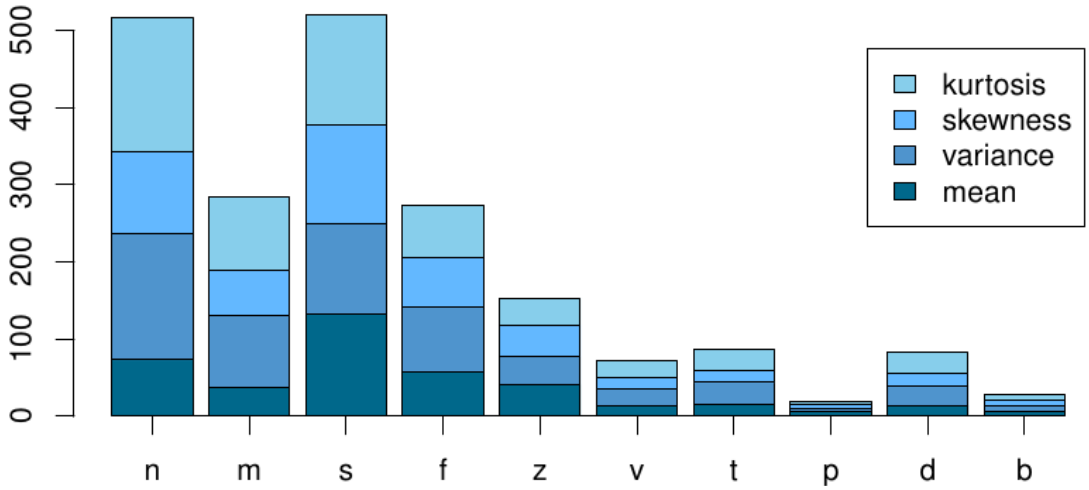

Figure 1 The F-ratios of the spectral moments of the labial and alveolar phonemes.

References:

Mook, Carola and Draxler, Christoph (2012) The speaker discriminating power of nasals, fricatives and vowels, Poster presented at the IAFPA 2012, Santander, Spain.

French Foreign Accent produced by Italians: The relation between effects of L1 prosodic background and L1 use

Rossana Cavone and Mariapaola D'Imperio

Aix-Marseille Université, CNRS, LPL, UMR 7309, Aix-en-Provence, France

The meaning of the term Foreign Accent (FA) *refers to the inability of a non-native speaker to reproduce the target language (L2) with the same precision as a native speaker* [1]. Past research has shown that the most important cues of FA are due to contact between the native and the non-native phonological and phonetic systems. During the contact phase, main features of the native-language (L1) remain underlying in L2 production as a result of transfer from L1 to L2.

The problem in this area of research is that nevertheless focused on the phonetic and phonological differences only in the field of segmental production [2], while prosody has been examined only in a relatively small number of studies [1-3-4]. Besides, there are also a few additional studies which attempt to elucidate the role of rhythm like cues of FA [5-6]. Moreover, many of the previous research cited, on second language acquisition, has focused on the prosodic properties, rather the indexical properties (or socio-phonetics) of the speech signal. However, a small number of studies have explicit examined the perception of indexical variation by second language learners but only at segmental level [7].

If we suppose, as [2] showed, that the development of new phonetic categories an L2 may be blocked by the continuative use of L1, giving origin to different degree of FA or inter-speaker differences, and if we consider that rhythm is part of the prosody of a language; foreign accent is not the result of intonational differences, but also a consequence of rhythmic errors and socio-phonetic factors.

For this reason, our priority is to underline the fact that influence of the speech rhythm on comprehensibility and perceived foreign accent was not distinguishable from the influence of other prosodic features and that the use of the L1 should be one of the cues of inter-speakers differences of FA degree.

I will outline discuss about specific questions of my research: Is the realization of different phonetic details of prosodic features that remains in the production of very advance L2 speakers? Is there a relationship between speaker-specific prosodic differences (FA degree) and the continued use of the mother tongue during the process of L2 learning?

References:

[1] Boula de Mareüil, P., Vieru-Dimulescu, B. (2006). The contribution of prosody to the perception of foreign accent, *Phonetica*, 63: 247–267.

[2] Flege, J.E., Frieda, E.M., Nozava, T. (1997). Amount of native-language (L1) affects the pronunciation of an L2, *Journal of Phonetics*, 25: 169–186.

[3] Mennen, I. (2004). Bi-directional interference in the intonation of Dutch speakers of Greek, *Journal of Phonetics*, 32(4): 543-563.

[4] Jilka M. (2000). The contribution of intonation to the perception of foreign accent, Thèse de doctorat, Université de Stuttgart, 2000.

[5] Munro, M.J. and Derwing, T.M. (2001). Modeling perceptions of the accentedness and comprehensibility of L2 speech - The Role of Speaking Rate*, Studies in Second Language Acquisition*, 23: 451-468.
[6] Tajima, K., Port, R.; Dalby, J. (1997). Effects of temporal correction on intelligibility of foreign-accented English, *Journal of Phonetics*, 25: 1–24.
[7] Clopper, C. G. , Bradlow, A. R. (2008). Perception of dialect variation in noise: Intelligibility and classification, *Language and Speech*, 51: 175–198.

/t/ realisation in RP English spontaneous speech

Emanuela Buizza

University of York

Voiceless alveolar plosives are known to be particularly prone to undergo connected speech processes in running speech in English (Cruttenden 2008). Several studies have focused on the assimilation, elision, tapping, voicing, glottalisation and frication of /t/ in varieties of English (see e.g. Foulkes and Docherty 1999, Harris 1990, Lavoie 2001). Despite the amount of research on Received Pronunciation (RP) English, the realisation of /t/ in this accent has not been investigated in detail yet.

This paper reports on a phonetic and phonological analysis of /t/ in a corpus of RP English spontaneous speech. For this study, a dataset of approximately 1000 tokens from 20 speakers of 'Modern' RP (Trudgill 2001) was collected through BBC radio podcasts. Auditory and acoustic analyses were carried out, including measurements of duration, amplitude and spectral moments analysis. Moreover, the phonological context of most frequent variants was also analysed.

The results of this study indicate that there is a wide range of variation in the realisation of /t/ in RP English too and that all speakers use a large set of variants. A major result is the identification of a significant number of fricated /t/s – realised without a complete closure – which raises the question of whether the contrast with the homorganic voiceless fricative /s/ is maintained. A comparison of fricated /t/s and phonological /s/s collected from the same corpus suggests that the two obstruents are likely to maintained their contrast. Moreover, the analysis of the distribution of the most frequent variants across segmental, syllabic and prosodic contexts and the analysis of inter- and intra-speaker variability confirm the great deal of variation in the realisation of /t/ in RP English spontaneous speech.

References:

Cruttenden, A. (2008) *Gimson's Pronunciation of English*. London: Hodder Education.

Foulkes, P. and Docherty, G.J. (1999) *Urban Voices: Accent Studies in the British Isles*. London: Arnold.

Harris, J. (1990) Segmental complexity and phonological government. *Phonology* 7, 255-300.

Lavoie, L.M. (2001) *Consonant Strength: Phonological Patterns and Phonetic Manifestations*. New York, London: Garland.

Trudgill, P. (2001) The sociolinguistics of modern RP. In Trudgill, P. (ed.) *Sociolinguistic Variation and Change*. Edinburgh: Edinburgh University Press, 171-180.

Language or Speaker? –

Investigating coronal-velar alternations in Vilela

Sven Grawunder

(Max Planck Institute for Evolutionary Anthropology, Dept. of Linguistics), Lucía Golluscio (CAICYT Universidad de Buenos Aires)

Scarcity of data is a common and, to some degree, growing problem with linguistic work on endangered languages. Especially phonetic questions about a speaker's specificity or a varieties particularity are hard to address, even when there are hours of speech with different sorts of text but only from one or two speakers available. This paper reports about such quest for disentangling the possible idiosyncrasy or systematicity, i.e ideolectal or language specifity of a single last speaker.

Fieldwork elicitation on the language Vilela (Lule-Vilela), a language of the Grand Chaco area, revealed an alternation between /t/ and /k/. For example we see in a set of suffixes (kanej-ki 'I'm tired'; tate-kis 'our father'), but even alternations within a single sequence of the same item do occur. However, earlier sources of the language from early 20th century back to missionary documents of the 18th century contain examples of similar fluctuation. Although it remains inevitably difficult to provide more data from other speakers of our target language Vilela, since it severely faces near extinction, we want to investigate the nature of this phenomenon and therefore we would nonetheless like to argue that our findings are a significant feature of Vilela and perhaps even part of an areal 'pattern'. Hence we separate the alternations out into two groups, those connected to a general tendency of assibilation involving coronalization and palatalization and those that are fluctuations between stops. The latter reminds of so-called "context-free change of place" (Blevins 2004, 122- 125) which has been discussed so far primarily for Austronesian (Blust 2004; Dunn 2006). In other words, we separate out production driven T>K and K>T fluctuations from more perception driven palatalization (e.g. k > ʧ) processes, which apparently both occur in our data. Perhaps a great part of these fluctuations are most likely to be recognized as "gestural intrusion errors". From recent speech error research (e.g. Marin et al. 2010) and speech acquisition research (Inkelas & Rose 2008) we know that also the opposite skewing (T>K) can result from intruding articulatory gestures. In our Vilela data we find these cases in repeated words containing stops of both places and within words with alternating syllable-initial stops. This corresponds to the conditions under which those speech errors usually are studied (Marin et al. 2010) so that we would like to apply such reasoning. For our speaker this is specific, since we are able to document fluctuations that occur in very short time spans and in elicitations, i.e. under conditions of stronger focus on the spoken form. Therefore we have to conclude further that here articulatory impediments (denture) as well as auditory impediments (age related hearing loss with depletion in high frequency space > 3kHz) come additionally as co-factors to those for speech planning into play.

The first pathway of our investigation is the specific nature of context, frequency and order of occurrence of the alternation that has been observed within nearly 10 years of

data collection. Apparently there are preferred contexts in terms of high vowels and a presumably typical velar-coronal order, which makes the phenomenon less context-free. Furthermore a general preference for coronal-velar shifts is demonstrated in the ASJP data base (Brown, Holman & Wichmann 2013), although the direction of the shift is not indicated. Chang et al. (2001) point out that, at least under experimental conditions (cf. Winitz et al. 1972), [ki] will be often confused as [ti], but [ti] will be hardly confused as [ki]. Plauché et al. (1997) had found that this confusal asymmetry is presumably facilitated by the high F2 of the adjacent high vowel, which causes the transitions of [ki] and [ti] to be similar. Additionally the pertained long VOT of the velar stop, including the higher amount of aspiration, would be re-analyzed by the listener as fricative part (e.g. [ʃ] or [ç]) of an affricate ([tʃ] or [tç] or the like). This describes the cases of coronalization, the latter including assibilation. Similarly would an assibilation of [th > ʧ, ts] work as reanalysis of an aspiration noise. On the other hand, a confused [ki] for [ti] would then necessarily be conditioned only by the transition information and require this high vowel context. Conversely, we see [t/k]-alternations also before /a/ and /e/ like in <kasé> and have therefore to assume that the transition of the second formant may not be an indispensable factor. Indeed, perceptual experiments using non-sense words show that burst information (durational and spectral) and vocalic transitions contribute and compete (cf. Neagu & Bailly 1997) for identification of place of articulation in stops. It had been shown for French that the burst information is capable to overwrite the information provided by the vowel transitions and that listeners are more likely to experience this with stops in contexts of high and mid front and high back vowels (Neagu & Bailly 1997, 1998). This provides a motivation for the observed sound change, but it also enables the reverse way of a [k > t]. Hence, we also report about a small-scale perception experiment based on the transcribed data with listeners from two linguistic backgrounds (Spanish and German). The results show that t/k-alternations do not predominantly occur due to a language bias of the transcriber although a difference for the two groups can be made.

Second, by means of the collected field data we show this alternation is salient in so far the transcribed categories find ground in acoustic parameterization. VOT is longest for the velar, but there are actually a number of others, including palatal and alveolar obstruents at the same level. Further we employed acoustic analysis of burst spectra using Discrete Cosine Transformation (DCT) and Amplitude difference (Ahi-Amid). These show an overlap of [t]- and [c]-like sounds in DCT2 and Ahi-Amid. The spectral curvature provides obviously a better means than the spectral slope in our data. Interestingly does the DCT1 not show a difference between [t] and [k]. And specifically the fact that those measures for [t] and [k] are the only non-significant result points towards the atypical [k]-spectrum and an unreliable spectral slope for these realisations. For burst intensity there appears only a significant difference between [ʧ] and [c], between [p] and [ʧ], and between [k] and [ʧ], a result that parallels that for the related DCT0 coefficient: only differences between [c] and [ʧ] and between [t] and [ʧ] are significant and support the stop (p, t, c) vs. affricate (ʧ) split. Since our speaker maintains the phonation contrast of marked/ejective stops (vs. unmarked/voiceless unaspirated) in a way that conflicts perceptually with place information cues, but feeds right into the pathway of assibilation as outcome of palatalization. So alveolar-stops can

merge with alveo-palatal affricates and velar ejective stops can become fronted, palatalized or palatal stops. The realization of glottalic stops as enforced aspirated or affricated stops fits well into this mechanic that Chang et al. (2001) had developed, at least for the alveolar and post-alveolar (palatal) stops.

References:

Blevins, J. (2004). Evolutionary Phonology: The Emergence of Sound Patterns. Cambridge University Press.

Blust, R. (2004). *t to k: An Austronesian sound change revisited. Oceanic linguistics, pages 365–410.

Brown, Cecil H., Eric W. Holman, and Søren Wichmann. In press (2013). Sound correspon-dences in the world's languages. Language 89.1

Chang, S., Plauché, M., and Ohala, J. (2001). Markedness and consonant confusion asymmetries. In Hume, E. and Johnson, K., editors, The role of speech perception in phonology, pages 79–101. Academic Press, San Diego, CA.

Inkelas, S. and Rose, Y. (2008). Positional neutralization: A case study from child language. Language, 83(4):707–736.

Marin, S., Pouplier, M., and Harrington, J. (2010). Acoustic consequences of articulatory variability during productions of/ t/ and/ k/ and its implications for speech error research. The Journal of the Acoustical Society of America, 127(1):445–461.

Neagu, A. and Bailly, G. (1997). Relative contributions of noise burst and vocalic transitions to the perceptual identification of stop consonants. In Kokkinakis, G., Fakotakis, N., and Dermatas, E., editors, Fifth European Conference on Speech Communication and Technology, EUROSPEECH 1997, Rhodes, Greece, September 22-25, 1997. ISCA.

Neagu, A. and Bailly, G. (1998). Cooperation and competition of burst and for- mant transitions for the perception and identification of french stops. In The 5th International Conference on Spoken Language Processing, Incorporating The 7th Australian International Speech Science and Technology Conference, Sydney Convention Centre, Sydney, Australia, 30th November - 4th December 1998. ISCA.

Plauché, M., Delogu, C., and Ohala, J. (1997). Asymmetries in consonant confusion. In Fifth European Conference on Speech Communication and Technology.

Winitz, H., Scheib, M., and Reeds, J. (1972). Identification of stops and vowels for the burst portion of /p, t, k/ isolated from conversational speech. The Journal of the Acoustical Society of America, 51(4B):1309–1317.

Speaker-specific behavior and acquisition

Alejandrina Cristia

MPI, Nijmegen

Becoming a native speaker and listener is an everyday miracle, particularly when we think of the rich variability in youngsters' everyday input. How do learners deal with speaker-specific behavior? How do they themselves come to contribute to this variation? In this talk, I will summarize three bodies of evidence from early first language acquisition which bear on these and other key questions of this summer school.

The first line of evidence relates to **how infants, toddlers, and children process the phonemic aspects of speech in the face of talker variation**. In broad terms, performance in learners of all ages is initially lowered when novel talkers are presented. Interestingly, even brief exposures to variable talkers (including not only different accents, but also very different voices) suffices to allow these young perceivers to regain their standing. The fact that these general stages are similar across individuals differing widely in age, and therefore linguistic competence, suggests that the perception system is astoundingly resilient. At the same time, differences across age groups underline the variety of strategies that perceivers could utilize in such situations.

While this productive research avenue assumes that learners must come to ignore variation to become native listeners, an emergent body of work begins to document **how children come to encode a range of variation, and actively use it**. Indeed, some evidence suggests that even 5-month-olds attend to a talker's accent when making social choices (such as accepting food from a stranger). Moreover, phonetic analyses reveal that, from their earliest productions, children are able to produce a range of phonetic targets that are necessary for them to convey their own social identity.

The final strand of work focuses not on how learners *in general* deal with variability, but rather on the **variation across different learners**. An increasing number of studies are documenting that stable variation in infant and toddler language and cognition reliably predicts their development later in life. The direct continuity and indirect cascades underlying the predictive value of early psycholinguistic measures is likely to contribute a key piece of evidence on our understanding of how children build language.

Children's use of prosody to compute syntactic structure on-line

Alex de Carvalho; Isabelle Dautriche; Anne Christophe

Ecole Normale Supérieure - Laboratoire de Sciences Cognitives et Psycholinguistique

Because the prosodic structure of an utterance tends to coincide with its syntactic structure, much previous work has suggested that children might use this cue to bootstrap syntactic acquisition (Christophe, Guasti, Nespor & van Ooyen, 2003; Morgan, 1996). Previous research has shown that adults exploit phonological phrase boundaries online to resolve syntactic ambiguities (Millotte et al., 2007;2008; Snedeker & Yuan, 2008; Kjelgaard & Speer, 1999; Schafer, 1997). However, with children, several studies have found no effect of prosody on their interpretations of structurally ambiguous sentences (Snedeker & Trueswell, 2001; Vogel & Raimy, 2002; Choi & Mazuka, 2003). To the best of our knowledge, only one recent study Snedeker & Yuan (2008) has found an effect of prosody on children's interpretation of globally ambiguous sentences. The effect, however, was rather weak, and children tended to stick to the interpretation they had heard first.

In order to pursue this question, we investigated whether young children are able to use the information provided by prosodic structure (i.e., phonological phrase boundaries) to constrain syntactic analysis. Pairs of homophones belonging to different syntactic categories (noun and verb) were used to create locally ambiguous sentences in French (e.g.a noun sentence: [*la petite* <u>*ferme*</u>] [est très jolie] – *the small farm* is very nice - vs. a verb sentence: [*la petite*] [<u>*ferme*</u> la fenêtre] – *the little girl closes* the window, where brackets indicate phonological phrase boundaries). Although both sentences start with the same three words, they can be disambiguated by the prosodic boundary that either directly precedes the critical word *ferme* when it is a verb or directly follows *ferme* when it is a noun. Crucially, all words following the homophone were masked, such that prosodic cues were the only disambiguating information.

In an oral completion task (Experiment 1), French 4.5-year-olds completed these three-word sentence onsets. Children successfully exploited prosody to interpret the ambiguous word as either a noun or a verb (similar to adults, Millotte et al., 2007).

To test whether children can use the information of prosodic boundaries online to conduct syntactic analysis, Experiment 2 used an intermodal preferential looking task with an eye-tracker. Children were presented with two images displayed side-by-side, one associated with the noun interpretation of the ambiguous word (e.g., a farm) and the other with the verb interpretation (e.g., a little girl closing something). We recorded both the time course of their eye-gaze and their pointing responses. Our results show that both 3.5- and 4.5-year-olds pointed toward the correct interpretation of the ambiguous word. Furthermore, the eye-tracking data show that children initially looked more toward the verb image (which always portrayed humans). In the noun sentence condition, children appropriately switched their gaze toward the noun image by the end of the ambiguous word, which, taking into account saccade preparation time, suggests that they computed its category before word offset.

Altogether, these results show that young children, upon hearing the first words of a sentence, exploit prosody on-line to group words into constituents, and exploit this constituent structure in their computation of syntactic categories.

Speaker and areal specific paths of acquiring neutralization of final stops in German

Sven Grawunder

Max Planck Institute for Evolutionary Anthropology, Dept. of Linguistics

Fuchs (2005: 173) states that neutralization is a process of articulatory reduction, allowing for speaker-specific "articulatory residues of the contrast". This paper seeks to explore the individual strategies that speakers develop and employ in reducing a contrast that is subject to a neutralization process, namely final devoicing of stops in German. This process is observable in cases like e.g. the derivation from plural forms into a singular form of a noun. Due to the process of final devoicing e.g. *Räder* 'wheels' - *Räte* 'councils' would share [ʁaːtʰ] 'weel/council' as a single phonetic output form (cf. Kohler, 1995; Wiese, 2000). A number of researchers (e.g. Charles-Luce 1985; Port and O'Dell, 1985; Piroth and Janker, 2004; Fuchs, 2005) found that these homophonic forms differ with regard to certain phonetic details like preceding vowel length, closure duration or burst duration.

This paper reports about a series of four experiments carried out using throughout ompatible settings. The list of auditorily presented stimuli consisted of pairs of artificial names (pseudo words) in plural forms following German phonotactics and differing in the medial stop, e.g. *Aus Salzburg kamen die Pru**g**e/Pru**k**e.* 'The Pruge/Pruke came from Salzburg.' Participants were asked to respond with a question that contained the singular form, e.g. *Wie sieht ein Prug/Pruk aus?* 'What does a Prug/Pruk look like?' This design (cf. Röttger et al. 2011) was meant to address some critique of previously applied methods. The factors addressed here additionally are mainly the areal bias of the subjects (cf. Piroth and Janker, 2004), adaptation towards the stimuli and the orthography bias, since one can not rule out that literate adults would envision how those pseudo-words would be written out (cf. Ziegler and Ferrand, 1998). Hence the primarily test variable over the series of experiments is group of participants (and its composition: age + area), besides the 'devoicing' of stops within. Moreover there are dialectal varieties of German, as the Alemannic ones, where especially velars are contrasting in voiceless stop (lenis) vs. affricate instead, whereas the areal standard would demand a stop neutralization at the same place. This offers a possibility to test influence of a competing glottalic pattern.

• Experiment 1: The participant group consisted of 12 young adults from the East Central Germany. The students were presented with 75 stimuli (50 critical/25 fillers) spoken by 3 Standard German speakers of the same area.

• Experiment 2: The participant group consisted of 13 pre-school children (ages 5,5 to 6,11) from the same area who were presented with 36 stimuli (no fillers) of the set for Experiment 1.

• Experiment 3: The participant group consisted of 9 young adults from Western Austria who were presented with the same 75 stimuli as in Experiment 1.

•Experiment 4: The participant group consisted of 10 young adults from Western Austria who were presented with the same 75 stimuli as in Experiment 1 but spoken by 2 speakers in Tyrolian dialect form.

All recordings with adults, using head-set microphone and EGG, were done under laboratory conditions. The audio-only recordings for Exp 2 were conducted under field conditions in a kindergarten.

The use of linear mixed effects models (LMMs) allowed to control for speakers and items as random factors so that overall effects (devoiced/neutralized vs. voiceless/non-neutralized stop) can be investigated, whereas generalized LMMs were use for testing individual acoustic and articulatory parameters as predictors. With inspection of speaker specific coefficients and effect sizes of differences between devoiced and voiceless stops we are able to explore and order trends within the group of participants. Preliminary results show significant overall effects for vowel duration, voicing-into-closure (accessed through closure velocity and band energy difference during closure) and f0 in Exp 1, but only for vowel duration (pMCMC=0.044) in Exp 2.

However, treating speakers as fixed effects we find usually one or two speakers per parameter falling out. Closer inspection of individual speakers reveals parameter expressions that are sometimes reversed to the overall trend. It provides for these cases evidence for trading relations since here e.g. F0 would be used instead of vowel duration. Although the results of Exp 1 parallel those of Röttger et al. (2011) with speakers from West Central Germany and Exp 3 parallel those of Exp 1, an areal specificity in the form of e.g. release realization for Tyrolian or closure velocity for East Central German. The strong tendency for quasi-unreleased 'devoiced' stops vis-á-vis aspirated stops in Exp 3 provides even a cross-linguistically often observable strategy for 'deaspiration' and thus neutralization. Especially the variation in voicing-into-closure sequences of the kids suggests a still ongoing acquisition of active devoicing mechanisms at this age, allowing still complete 'maintenance' of the contrast.

Speech errors in Exp 2 also indicate a strong influence of similar sounding neighbors of the lexicon. These slips into real words of German show even independence of word class and syntactic slot. Another interesting observation is that for some speakers, adults and pre-schoolers, the onset of words (#CV/#CCV) varies with respect to the final devoicing pattern. This is reminiscent of findings of onset differences in minimal pairs with coda contrast by Hawkins and Nguyen (2005) but offers a hint on more effortful planning which could at the end be caused by the task itself.

References:

Charles-Luce, J. (1985). Word-final devoicing in German and the effects of phonetic and sentential contexts. The Journal of the Acoustical Society of America, 77(13):309–324.

Fuchs, S. (2005). Articulatory correlates of the voicing contrast in alveolar obstruent production in German. PhD thesis, Queen Margaret University College published in ZASPIL series no 41.

Kohler, K. (1995). *Einführung in die Phonetik des Deutschen*. 2nd edition, E. Schmidt.

Piroth, H. and Janker, P. (2004). Speaker-dependent differences in voicing and devoicing of German obstruents. *Journal of Phonetics*, 32(1):81–109.

Port, R. F. and O'Dell, M. L. (1985). Neutralization of syllable-final voicing in German. *Journal of Phonetics*, 13(4):455–471.

Röttger, T., Winter, B., and Grawunder, S. (2011). The Robustness of Incomplete Neutralization in German. *Proceedings of the XVIIth ICPhS, 17-21 August 2011 HongKong*, 1722–1725.
Wiese, R. (2000). *The phonology of German*. Oxford University Press, USA.
Ziegler, J. C., and Ferrand, L. (1998). Orthography shapes the perception of speech: The consistency effect in auditory word recognition. Psychonomic Bulletin and Review, 5, 683–689.

The influence of second language on speaker-idiosyncratic temporal patterns

Marie-José Kolly, Volker Dellwo, Adrian Leemann

Phonetics Laboratory, University of Zurich
marie-jose.kolly@pholab.uzh.ch, volker.dellwo@uzh.ch,
adrian.leemann@pholab.uzh.ch

Can we determine speaker identity based on the acoustic signal when speakers talk in different languages? It has previously been suggested that time domain measurements are robust to within-speaker variation: For example, the percentage over which speech is vocalic varies considerably between different speakers (Wiget et al., 2010) and there is first evidence that this is unaffected by speech rate (cf. Dellwo/Koreman, 2008, Dellwo et al., forthcoming), by dialect disguise (Dellwo et al. 2009), and by speaking style (Leemann et al., submitted). However, the degree to which this variability is affected by the speakers' choice of language is unclear. Or, in other words, do speakers transfer idiosyncratic time-domain features from their L1 to an L2?

From a phonetic/phonological perspective there are multiple sources of influence on L2 speech temporal characteristics that are, for example, (a) carried over from the speaker's L1 to the L2, such as voice onset time (cf. Bohn 1998) or (b) reflected by general properties of L2 speech that are independent of a specific L1 (L2 speakers tend to lengthen vowels in a foreign language; Adams/Munro, 1978). It seems plausible that speaker-idiosyncratic influences also exist. For example, two similarly proficient L2 speakers of the same L1 might have different habits or preferences concerning the realization of temporal characteristics in the L2. In the present research we are testing the hypotheses whether (a) speakers show idiosyncratic temporal characteristics in L2 speech and, if yes, whether (b) the speaker-idiosyncratic differences are independent of the performed language (L1 or L2).

We are currently examining 16 sentences of each Zurich German (L1) speech and French and English (L2) speech of 16 speakers (8f, 8m). Automatic durational measurements such as the percentage over which speech is voiced or the standard deviation of voiced or voiceless intervals (Dellwo et al., 2012, Leemann et al., submitted) are calculated for each sentence. First results reveal support for the two hypotheses.

References:

Adams, C./Munro, R. R. (1978): In search of the acoustic correlates of stress: Fundamental frequency, amplitude and duration in the connected utterance of some native and non-native speakers of English. *Phonetica* 35: 125–156.

Bohn, O.-S. (1998): Wahrnehmung fremdsprachlicher Laute. Wo ist das Problem?. In: Wegener, H. (ed.): *Eine zweite Sprache lernen. Empirische Untersuchungen zum Zweitspracherwerb*. Tübingen, Narr: 1–20.

Dellwo, V./Koreman, J. (2008): How speaker idiosyncratic is measurable speech rhythm? Abstract presented at the annual IAFPA meeting 2008, Lausanne/Switzerland.

Dellwo, V./Ramyead, S./Dankovicova, J. (2009): The influence of voice disguise on temporal characteristics of speech. Abstract presented at the annual IAFPA meeting 2009.

Dellwo, V./Leemann, A./Kolly, M.-J. (2012): Speaker idiosyncratic rhythmic features in the speech signal. Proceedings of Interspeech 2012, 9.–13.9.2012, Portland (OR), USA.

Dellwo, V./Leemann, A./Kolly, M.-J. (forthcoming): Rhythmic variability between speakers: Articulatory and linguistic factors.
Leemann, A./Dellwo, V./Kolly, M.-J. (submitted): Speaker discrimination based on speech temporal features.

Biomechanics

Pascal Perrier

Gipsa-lab – CNRS – Université Grenoble-Alpes

At the intersection of life sciences, mechanics and mathematics, biomechanics investigates and models the mechanical structures and the mechanical properties of the human body and their influences on human movements. Biomechanical properties include muscle force generation mechanisms, muscle force directions, joints characteristics, soft tissue stiffness and soft tissue deformation principles. As a consequence they are strongly dependent on the morphology of the body. A number of motor control studies based on biomechanical models of the arm, the legs and, for speech, the orofacial articulators have shown that biomechanical factors contribute to shape movement trajectories, influence gestural accuracy, and are in turn taken into account in the specification of the motor control strategies underlying movement generation (see among others Flanagan et al., 1993, Franklin et al, 2007, Gomi & Kawato, 1996 versus Gribble et al, 1998, Nazari et al., 2011, Perrier et al., 2003, Perrier & Fuchs, 2008, Pinter et al., 2012). Differences between female and male voices associated with differences in vocal folds mechanical properties are a very obvious example of these potential influences.

As a consequence it is expected that inter-subject variability in body size, body morphology, or muscle force generation capacities should significantly influence movements. In studies of arm or legs motor control, inter-subject variability has been considered. Interestingly, this was mainly done in order to show that the Central Nervous System is to a certain extent able to overcome it: general motor control strategies apply to all subjects independently. It was thus shown that during backward giant swings on the high bar experienced gymnasts use strongly similar neuromuscular control strategies of the upper-limb and trunk muscles in spite of their morphological differences (Frère & Hug, 2012). Hug et al. (2010) observed a certain amount of variability in the EMG patterns of the lower limb of trained cyclists during pedaling, but they also showed that these patterns were underlying the same number of very similar muscle synergies. Chu et al. (2013) investigated movement variability in children with secondary dystonia. Dystonia is a motor disorder associated with movements that are slow and inaccurate. Chu et al (2013) have shown that children with dystonia were able to deal with the pathological specificities of their motor apparatus and displayed patterns of variability that were very similar to those of healthy children. Similarly, speech motor control studies usually observe cohorts of subjects in order to discuss common motor control strategies. In these cases, the inter-subject variability is essentially ignored.

In opposition to this approach, we will focus in the presentation on inter-speaker variability in speech production and on its potential links with biomechanical characteristics of the oro-facial speech production system. To do so we will first present data from the literature and analyze how the observed variability informs about biomechanical factors or can be linked to them (see for example Gomez-Vilda et al.,

2013). Second, using a biomechanical model of the jaw-tongue system we will analyze how biomechanical factors such as vocal tract and head morphology, muscle force directions and soft tissue stiffness can affect speech movement amplitudes, trajectories, and accuracy.

References

Flanagan, J.R., Ostry, D.J, & Feldman, A.G. (1993). Control of trajectory modifications in target-directed reaching. *Journal of Motor Behavior*, 25(3), 140-152.

Franklin, D.W., Liaw, G., Milner, T.E., Osu, R., Burdet, E., & Kawato, M. (2007). Endpoint stiffness of the arm is directionally tuned to instability in the environment. *Journal of Neuroscience*, 27(29), 7705-7716.

Frère, J. & Hug, F. (2012). Between-subject variability of muscle synergies during a complex motor skill. *Frontiers in Computational Neuroscience*, Vol.6, Paper 99 (13 pages)

Gómez-Vilda, P., Rodellar-Biarge, V., Nieto-Lluis, V., Muñoz-Mulas, C., Mazaira-Fernández, L.M., Martínez-Olalla, R., Álvarez-Marquina, A., Ramírez-Calvo, C., & Fernández-Fernández, M. (2013). Characterizing neurological disease from voice quality biomechanical analysis. *Cognitive Computation* (On Line February 2013)

Gomi, H., & Kawato, M. (1996). Equilibrium-point control hypothesis examined by measured arm stiffness during multijoint movement. *Science*, 272(5258), 117-120.

Gribble, P.L., Ostry, D.J., Sanguineti, V., & Laboissière, R. (1998). Are complex control signals required for human arm movement? *Journal of Neurophysiology*, 79(3), 1409-1424.

Hug, F., Turpin, N.A., Guével, A. & Dorel, S. (2010). Is interindividual variability of EMG patterns in trained cyclists related to different muscle synergies? *Journal of Applied Physiology*, 108, 1727-1736.

Nazari, M.A., Perrier, P., Chabanas, M., & Payan, Y. (2011). Shaping by stiffening: a modeling study for lips. *Motor control*, 15(1), 141-168.

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.

Perrier, P., & Fuchs, S. (2008). Speed-curvature relations in speech production challenge the 1/3 power law. *Journal of Neurophysiology*, 100(3), 1171-1183.

Pinter, I.J., van Soest, A.J., Bobbert, M.F., & Smeets, J.B.J. (2012). Conclusions on motor control depend on the type of model used to represent the periphery. *Biological Cybernetics*, 106, 441–451

Speaker-specific behavior of 4-year-old Canadian French children: ultrasound data analysis in the light of biomechanical knowledge

Guillaume Barbier[1], Pascal Perrier[1], Lucie Ménard[2], Yohan Payan[3], Mark K. Tiede[4], Joseph S. Perkell[5]

[1] Speech and Cognition Department, GIPSA-lab & Grenoble University, Grenoble, France
[2] Department of Linguistics, Université du Québec à Montréal, Montréal, Québec, Canada
[3] TIMC-IMAG Laboratory, Grenoble University, Grenoble, France
[4] Haskins Laboratories, New Haven, Connecticut, USA
[5] Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

Introduction

This work is based on the hypothesis that speaker-specific articulatory strategies, aiming at reaching desired acoustic characteristics, are influenced by vocal tract morphological and biomechanical properties. In turn, these articulatory strategies are likely to influence the emergence of perceptual phonetic categories in the context of perception-action interaction. In an ontogenetic perspective, we speculate that this stage of speech production development could largely influence the phonetic organization in adult speech communication. Using experimental data and biomechanical models of children's vocal tracts, this work contributes to the understanding of speaker-specific articulatory and motor behavior of 4-year-old Canadian French children.

Methods

Data: We designed a study [Barbier et al., Interspeech 2013] in which we recorded acoustic and ultrasound data from twenty 4-year-old children and from ten young adults for comparison. Participants had to recall, plan and pronounce utterances in an experiment designed as a puppet game (puppets names are isolated vowels and V_1-C-V_2 sequences). In this study, we use the HOCUS system [Whalen et al., JSLHR 2005], which uses optical tracking (Optotrak, NDI Certus) of infrared emitting diodes (iREDs), positioned both on the ultrasound probe and on the head of the participant to provide a representation of the data in a movement-corrected, head-centric frame of reference.

Modeling: The existing 2D biomechanical model of the tongue [Payan and Perrier, Speech Comm.1997] and jaw [Zandipour, PhD Thesis, Boston U., 2006] was adapted to 4-year-old children morphologies thanks to a method proposed by [Winkler et al., Interspeech 2011]. The speaker-specific adaptation is made thanks to sagittal cephalometric data (eight 4-year-old speakers thus far). These adapted models give a fair account of speaker-specific vocal tract morphologies and anatomical muscles implementations and give a globally realistic description of the physical properties of the speech articulators. These models are controlled by muscular activations.

Expected Results

Articulatory and acoustic inter-speaker and intra-speaker variability were assessed for 5 vowels (/i e ɛ a u/) and 3 voiced stops (/b d g/), for 20 children and 10 young adults. As

expected, intra-speaker variability is significantly greater in children than in adults. Currently, this work aims at inferring motor commands patterns through the use of the models, these latter enabling the reproduction of the speaker-specific tongue contours. The motor commands patterns, together with acoustic and articulatory characteristics, will allow us to construe speaker-specific variability, in children and in the differences between adults and children.

Acknowledgements: This work is partly supported by FQRNT project N° 147877 (Lucie Ménard) and the ANR project ANR-08-BLAN-0272 (Pascal Perrier). We want to thank Amélie Prémont, Christine Turgeon, Paméla Trudeau-Fisette and Dominique Côté for their crucial help concerning the experiment. We are thankful to the infant school Le Sablier and particularly its co-director Carole Langlois for her involvement in the project. We are also very grateful to all our participants.

Contributions of the vocal tract shape to the C-center effect

Jana Brunner, Christian Geng, Adamantios Gafos

Universität Potsdam

Previous work has suggested that there are differences in the temporal organization of CV and CCV syllables. In our recent study on German, for example, we show that in syllables starting with a cluster, such as *queren*, the prevocalic consonant tends to be closer to the vowel than in some syllables starting with a singleton consonant, such as *wehren*. This results in a shorter vowel duration in CCV as compared to CV syllables, an effect which has been termed *C-center effect* because what is temporally stable across CCV and CV syllables is the interval from the consonantal center (the middle of the cluster or the singleton) to the end of the vowel.

Most studies argue that the C-center effect emerges from competitive coupling of articulatory gestures in the onset. The aim of the present study is to test an alternative or additional explanation for the effect, i.e. whether the vocal tract shape influences gestural timing. In our previous work we suggest that the C-center effect is largely grounded in coarticulation. Coming back to the example *queren-wehren*, due to carry-over coarticulation with the velar the tongue is higher during the /v/ in *queren* than during the /v/ in *wehren*. As a result of that, the distance the tongue has to travel to come from /v/ to /eː/ in *queren* is shorter than the distance from /v/ to /eː/ in *wehren*. For the shorter movement in *queren* the tongue needs less time than for the larger movement in *wehren*. This results in the C-center effect: the vowel duration, measured as the interval from the end of prevocalic /v/ to the end of the vowel is longer in *wehren* than in *queren*. Our data also show that in items with diphthongs containing low vowels (e.g. *gleiten* versus *leiten*) no C-center effect is observed. This is because in these items the tongue is again higher in cluster-/l/ than in singleton-/l/, but since the vowel is low, the distance the tongue has to travel to reach the vowel following the cluster is greater in *gleiten* than in *leiten*. Consequently, the vowel in the CCV item is not shorter, but often longer than the vowel in the CV item and usually no C-center effect is observed.

Since the distance the tongue has to travel depends, among other things, on the speaker's vocal tract shape, there should be a relation between vocal tract shape and the occurrence of a C-center effect. Speakers differ in the ratio of their oral versus pharyngeal tube lengths. Whereas males tend to have a proportionally longer pharyngeal tube, females tend to have a proportionally longer oral tube. Related to that, males tend to rely on varying tongue height in order to produce phonemic differences, and females tend to rely on varying the horizontal tongue position. Great differences in tongue height in *gleiten* versus *leiten* should therefore be typical for speakers with variability in the vertical dimension (i.e. predominantly male speakers), and there should be less of an influence for speakers varying the horizontal dimension. The hypothesis of the present investigation is thus that speakers with more variation in the vertical dimension will have negative vowel compression (i.e. a longer vowel in *gleiten* than in *leiten*), whereas speakers with variability in the horizontal dimension will tend

to have equal vowel durations or maybe even a shorter vowel in *gleiten* than in *leiten* (C-center effect).

Five speakers (two males, three females) were recorded via EMA speaking items with /gl/ and /l/ onsets followed by low vowels. In order to assess whether those speakers vary rather in the horizontal or the vertical dimension the variability of the tongue tip sensor along the palatal outline was measured. Afterwards, the variability of the tongue tip sensor perpendicular to that dimension was measured. Finally, a ratio of the two variabilities was calculated. The speakers' ratios varied continuously, with females towards the upper end and males towards the lower end. Articulatory movements were segmented. For each item the temporal distance from the end of the prevocalic consonant /l/ to the consonant following the vowel (i.e. /t/ in *gleiten*) was measured. Vowel compression was measured as the difference in that interval between the CCV and the CV item. Positive vowel compression values mean that the cluster vowel is shorter than the singleton vowel.

Spearman correlations between ratios and vowel compression were calculated. The results show that, in line with the hypothesis, for five of six item pairs there is a significant positive correlation: speakers with high ratios (more horizontal variability) have vowel compression values above 0. Speakers with low ratios (more vertical variability) have high negative vowel compression values. Thus, speakers with more horizontal variability (predominantly females) are more likely to produce a C-center effect in /gl/ clusters with low vowels than speakers with more vertical variability (predominantly males).

New methods to explore the complexity of the behavior of the larynx and its interaction with breathing factors and configuration of the tongue.

Leonardo Lancia and Sven Grawunder

Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

We propose some new method which permits to better observe the complex functioning of the larynx and its interactions with factors affected by other anatomical structures (like subglottal pressure and tongue position). These methods are applied to analyze data from a speech production experiment in which 5 German female participants aged between 20 and 30 were asked to utter repeatedly and without interruption during 10 seconds dysillabic utterances of the kind $C_1VC_2C_1V$ or VC_1C_2V. In these utterances the C_2 was always /s/, the C_1s could be either both /k/ or both /t/, and the Vs could be either both /a/ or both /i/. This experimental design permitted the independent manipulation of the following factors: 1) presence versus absence of glottalization in intervocalic position (expected to affect the first vowel of an utterance much more frequently when the utterance starts with a vowel); 2) stress on the vowel preceding the glottalization or on the vowel following glottalization (stress on the 2nd syllable vs. stress on the 1st syllable on utterances starting with a vowel); 3) fronting and raising of the tongue during the production of the vowels (fronted and raised during /i/ but retracted ad lowered during /a/); 4) fronting of the tongue during the production of the consonants (fronted in /t/ but retracted during /k/) and 5) available volume of air (more air at the beginning of the sequence than at the end, since speakers were asked not to breath during a sequence of repeated utterances). We simultaneously recorded the acoustic signal, ultrasound video showing the movement of the back of the tongue and the electro-glottographic signal (henceforth EGG) which permits to monitor the activity of the vocal folds. The analysis focused on the vocalic portions of the signals. The acoustic signal was used to extract intensity, F0, F1 and F2 curves. We used a variant of recurrence analysis recently proposed by Lancia, Avelino and Voigt (submitted) to extract from the EGG signal a time varying measure of the regularity of the shape of vocal folds' vibrations. Also the closing quotient was extracted from each cycle of vocal fold's vibration. Wavelet based Functional Mixed Models (henceforth WFMM, Morriss and Carrol, 2006) were used to assess the statistical significance of the effects of the experimental factors (vowel quality, consonantal context, presence of glottalization etc.) on the evolution over time of the measured quantities. Analyses conducted so far show complex interactions of the experimental factors on the evolution over time of almost all the recorded parameters. For example, as expected, the F0 is higher for /i/ than for /a/ and higher in stressed vs unstressed position. However this happens when the vowel is not preceded by glottalization. Also the nature of the vowel produced shows a significant effect on the amount of regularity of the vocal fold vibration, providing new evidence for the link between tongue position and vocal folds vibration. Since the WFMM approach can also be applied to model changes in images, we will use this method to model the effect of the variation in the acoustic and voice quality parameters on the speaker specific

configurations of the tongue. The discussion will focus on the changes of rate and quality of vocal folds vibration and their relation to the configuration of the tongue.

Possibilities and limits of speaker-specific articulatory variability

Melanie Weirich

Friedrich-Schiller-Universität Jena, Germany

"Idiosyncratic features of a person's speech" may "be a part of an individual's learned speech behaviour" or due to "anatomical and physiological considerations" (Ladefoged & Broadbent 1957, p. 98). Sociolinguistic studies focus on the learned external non-biological influences on speech and the speaker's (conscious or unconscious) choice to develop an individual speech style that mirrors "local history and personal desire" (Foulkes & Docherty 2006, p. 85). However, idiosyncratic choices in speech production also have limits. Physiological and biomechanical parameters such as larynx morphology, vocal tract geometry, palatal shape, and tongue muscles affect the speech production process in general (Stone 1991, Vorperian & Kent 2007). Most interestingly, individual differences in these biological determinants confine speaker-specific strategies used in sound production (Weirich & Fuchs acc., Weirich et al. subm., Weirich 2012, Brunner et al. 2009). Studying these individual differences in more detail can help us understand how the production and coarticulation of speech sounds works. More specifically, it can facilitate our understanding of the possibilities and limits of inter-speaker variability also with respect to the phoneme inventory of a language and its phonemic contrasts.

In this talk I will focus on three studies investigating inter-speaker articulatory variability in German by means of Electromagnetic Articulography. While two of these investigations deal with the speech of monozygotic and dizygotic twin pairs and thus focusing on the relationship between physiological similarity and inter-speaker variability, the third study highlights differences in male and female speech, which have been explained both in terms of biophysical inevitabilities (e.g. Fant 1966, Simpson 2001, Winkler et al. 2006) as well as learned behaviours (e.g. Henton & Bladon, 1988, Labov 1990). The speech material analyzed comprises the realization of the phoneme contrast /s/ - /S/, the looping trajectories of the tongue dorsum during vowel-stop-vowel sequences, and the sizes and shapes of articulatory vowel spaces. These studies highlight not only the significance of investigating speaker-specific variability but also its interrelation with the dynamic nature of speech.

References:

Brunner, J., Fuchs, S. & Perrier, P. (2009) On the relation of palate shape and articulatory behaviour. *Journal of the Acoustical Society of America*, 125, 3936–3949.

Fant, G. (1966) A note on vocal tract size factors and non-uniform F-pattern scaling, STL-QPSR 4, 22–30.

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

Henton, C. G. & R. A. W. Bladon (1988) Creak as a sociophonetic marker, In Larry M. Hyman and Charles N. Li (eds.), *Language, Speech and Mind: Studies in Honour of Victoria A. Fromkin*. London: Routledge, 3–29.

Labov, W. (1990) The intersection of sex and social class in the course of linguistic change, *Language Variation and Change* 2, 205–254.

Ladefoged, P. & Broadbent, D. (1957) Information conveyed by vowels. *Journal of the Acoustical Society of America* 29(1), 98-104.

Simpson, A. P. (2001) Dynamic consequences of differences in male and female vocal tract dimensions, *Journal of the Acoustical Society of America* 109, 2153–2164.

Stone, M. (1991). Toward a model of three dimensional tongue movements. *Journal of Phonetics*, 19, 309–320.
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.
Weirich, M. (2012) *The influence of NATURE and NURTURE on speaker-specific parameters in twins' speech: Articulation, acoustics and perception*. PhD thesis. Humboldt-Universität zu Berlin.
Weirich, M. & Fuchs, S. (acc.) Palatal morphology can influence speaker-specific realizations of phonemic contrasts. *Journal of Speech, Language and Hearing Research*.
Weirich, M., Lancia, L. & Brunner, J. (under review) Inter-speaker articulatory variability during vowel-consonant-vowel sequences in twins and unrelated speakers. *Journal of the Acoustical Society of America*.
Winkler, R., Fuchs, S. & Perrier, P. (2006) The relation between differences in vocal tract geometry and articulatory control strategies in the production of French vowels: Evidence from MRI and modelling. *Proceedings of the 7th International Seminar on Speech Production*, Ubatuba, 509-516.

Speaker-specific breathing profiles during spontaneous speech

Amélie Rochet-Capellan[1] *and Susanne Fuchs*[2]

[1] CNRS, Gipsa-lab, Département Parole et Cognition, Grenoble
[2] Zentrum für Allgemeine Sprachwissenschaft, Berlin
Amelie.rochet-capellan@gipsa-lab.grenoble-inp.fr, fuchs@zas.gwz-berlin.de

”Speech is organized in terms of the expiratory air flow from the lung” (Lieberman, 1967, p.52). This organization defines the breath group that corresponds to the chunk of speech produced on a single exhalation. Previous studies have found that in both read and spontaneous speech, the properties of the breath group, and their relations to inhalation parameters, are speaker-specific [1, 2] and varied with age [3] or cognitive load [4]. We previously analyzed the linguistic structure of the breath group in spontaneous speech produced by 26 native female speakers of German [5]. We found that most of the breath groups included 1-3 clauses. In average, groups with 1-3 clauses lasted 3.5 s and included 16 syllables. Half of these groups started with a matrix clause; a quarter with an embedded clause and the last quarter with an incomplete clause (continuation, repetition, hesitation). We also found that the amplitude and the duration of the inhalation varied as a function of the first clause type and with respect to breath group length. Breath groups including vocalized hesitations also coincide with deeper inhalation. These average results demonstrate the interplay of speech-planning and breathing control in spontaneous speech: to some extent the linguistic structure of the breath group is anticipated during the inhalation phase. We also found large variability between speakers. This variability was found in the number of breath groups produced by each speaker. On average subjects realized 1 to 10 breath groups to summarize a short text. Speaker-specificity was also evident in the proportion of breath groups with hesitation, ranging from 0 to more than 50% according to the speaker. In this presentation, we will discuss in more details the characteristic of the breath group with respect to the speaker and discuss it in light of speech planning strategies.

References:

[1] Teston, B. and Autesserre, D. (1987). L' aérodynamique du souffle phonatoire utilisé dans la lecture d'un texte en français. in International Congress of Phonetic Sciences (ICPhS). Estonia, University of Tallin. p. 33-36.
[2] Fuchs, S., Petrone, C. Krivokapic, J. & Hoole, P. (2013). Acoustic and respiratory evidence for utterance planning in German. Journal of Phonetics 41. 29-47.
[3] Sperry, E. E. & Klich, R. J. (1992). Speech breathing in senescent and younger women during oral reading. J Speech Hear Res, 35, 1246-1255.
[4] Mitchell, H. L., Hoit, J. D., & Watson, P. J. (1996). Cognitive-linguistic demands and speech breathing. J Speech Hear Res, 39, 93-104.
[5] Rochet-Capellan et Fuchs (Submitted). The interplay of linguistic structure and breathing in German spontaneous speech.

An Acoustic Study of Sustained Vowels produced by Patients with or without Recurrent Laryngeal Nerve Paralysis after Thyroid Surgery

Camille Fauth[1], Béatrice Vaxelaire[2], Jean-François Rodier[3], Pierre-Philippe Volkmar[3], Marion Béchet[2], Rudolph Sock[2]

[1] LORIA CNRS UMR 7503 - Nancy
[2] Université de Strasbourg, Institut de Phonétique de Strasbourg – France – IPS & U.R. 1339 Linguistique, Langues et Parole – LilPa, E.R. Parole et Cognition
[3] Centre Paul Strauss – Strasbourg – France, Département de Chirurgie Oncologique

The aim of this study is to assess the consequences of thyroid surgery on the voice of patients, in order to identify various perturbations which this surgery may provoke, and also to reveal possible compensatory strategies or readjustments that the patient may develop, alone or with speech therapy; this is a longitudinal study.

The assessment was based on the analysis of acoustic signals, from which a large amount of cues related to voice quality, but also to articulatory behaviour of speakers was extracted. Production of sustained vowels allowed carrying out two spectral studies. *Measurements* obtained, using the speech editor PRAAT© (window = 0,025s), for the three extreme vowels were: F0 (Hz); Harmonics-to-Noise Ratio or HNR (dB); F1 (Hz) and F2 (Hz); and Vowel Space Area or VSA (kHz^2). The VSA, or area of triangles, was calculated using Heron's formula.

The first experiment deals with the spectral characteristics of the voice of 7 patients who underwent thyroidectomy, with no laryngeal paralysis. The data were obtained in 3 phases: 1) a pre-surgery phase, the day before surgery, which gives us the patient's unaltered *reference* voice; 2) a first post-surgery phase, the day after surgery, when the patient's voice is altered in varying degrees; 3) a second post-surgery phase, 15 days after surgery, which allows observing probable voice and speech recuperation. The second study was conducted based on spectral characteristics of the voice of 7 speakers with laryngeal post-thyroidectomy paralysis. The experimenter had unfortunately not been able to collect data from the patients before surgery (due to hospital and admittance constraints). First recordings could only be carried out in the second post-surgery phase, *i.e.* 15 days after surgery. Thus in cases when a patient had only been recorded in post-surgery phases, data were acquired from a healthy control speaker, matched with age and gender. Such data served as reference for pathological subjects. All patients were seen and recorded once a month during their speech therapy sessions.

This investigation has allowed us to verify the relevance of the measured acoustic parameters in evaluating voice alterations due to thyroidectomy, with or without unilateral paralysis. We have also gained insights into the capabilities of patients in deploying strategies for voice recovery after surgery, with or without speech therapy. Operation of the larynx appears to have consequences, at least in the short term, on the patient's voice.

This study also revealed compensation strategies that speakers are able to develop with speech therapy. These include an awareness of the oro-laryngeal gestures that help acquire an acceptable voice quality, presumably also to self-correction by auditory feedback.

The concept of a "target" as a control space for execution of possible articulatory and acoustic entities, which are perceptually acceptable, seems particularly relevant in this study, since disordered speakers reorganise their productions according to their own physiological and anatomical constraints, caused by the disorder. Note that speaker variability is still very remarkable in postoperative recording phases. It might be useful to increase the number of speakers in order to classify patients into subgroups according to the impact of surgery on the patient's voice. Additional recordings are also underway to increase the number of subjects for each group.

References:

Fauth C, Vaxelaire B, Rodier J, Volkmar P, Bouarourou F, Hirsch F, et al.: A spatiotemporal prospective study of speech in patients with or without recurrent laryngeal nerve paralysis after thyroid surgeryMontréal, 2011, p 57-64.

Laccourreye O, Malinvaud D, Ménard M, Bonfils P: Immobilités laryngées unilatérales après chirurgie de la glande thyroïde. Journal de Chirurgie 2009;146:553-558.

Laccourreye O, El Sharkewy L, Hans S: Paralysie laryngée unilatérale après chirurgie de la glande thyroïde. La lettre d'Oto-rhinon-laryngologie et de chirurgie cervico-faciale 2005;:12-16.

Mohil RS, Desai P, Narayan N, Sahoo M, Bhatnagar D, Venkatachalam VP: Recurrent laryngeal nerve and voice preservation: routine identification and appropriate assessment - two important steps in thyroid surgery. Ann R Coll Surg Engl 2011;93:49-53.

Laccourreye O, Le Clech G, Santini J: Complications de la chirurgie des affections et tumeurs bénignes de la thyroïde; in : Les risques chirurgicaux en oto-rhino-laryngologie : information, prise en charge et prévention, Société francaise d'oto-rhino-laryngologie. Paris, Laccourreye O, Chabolle F., 2008, p 369-386.

Trésallet C, Chigot J-P, Menegaux F: How to prevent recurrent nerve palsy during thyroid surgery? Ann Chir 2006 févr;131:149-153.

Generalized Linear Mixed Models and how to run them in R

Roger Mundry,

Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

Differences between speakers are commonplace in linguistic data. Hence, if multiple data points from the same respective speakers are included in a statistical analysis this needs to be accounted for because otherwise the data would be 'pseudo-replicated' leading to an inflated probability of false positive significances. Generalized Linear Mixed Models (GLMM) provide a powerful and very flexible tool for the statistical analysis of data comprising repeated data from the same speakers. Besides allowing for effectively controlling the error rate, they also allow to gain insights into differences between speakers. For instance, speakers could just differ in an average value but could also differ with regard to the difference they make, for instance, between two vowels. Mixed Models allow to measure and control for both such components of speaker differences. In the first section of the course I shall give a brief introduction to GLMMs and other statistical concepts needed to use them appropriately (e.g., interactions or dummy coding). In the second, I shall give an introduction to how to run mixed models in R. R is a very powerful and flexible statistical software and makes GLMMs readily available. Participants will get hands on experience in using R for running GLMMs, checking whether their assumptions are fulfilled and exploring their results.