British Association of Academic Phoneticians

2016 Colloquium Lancaster University

Book of abstracts

The effect of visual speaker gender on the perception of high back vowel fronting in SSBE

Roy Alderton (Lancaster University)

Research has found that the high back vowels /uː/ and /ʊ/ of the GOOSE and FOOT lexical sets are being increasingly realised in fronted positions in many varieties of English around the world, including Southern Standard British English (SSBE), and that this change is being led by young women in terms of production (e.g. Williams & Kerswill, 1999; Fabricius, 2007). Following the exemplar model of speech perception, some studies suggest that fronted GOOSE and FOOT may differ from other phonetic variables by not indexing any social characteristics in terms of perception (Fridland, 2012; Haddican *et al.*, 2013), which indicates that this sound change is not perceptually associated with the speakers who are said to be most advanced in production. Other investigations, however, have found that fronted variants of these vowels may be perceptually linked with younger speakers in SSBE (Harrington *et al.*, 2007, 2008; Kleber *et al.*, 2012).

I describe a study that uses a speech perception experiment to investigate whether the gender of the speaker affects perception of fronted GOOSE and FOOT in SSBE in a similar way to age, reflecting findings stating that young females produce more fronted variants. Participants completed a binary forced-choice word identification task based on continua of digitally manipulated realisations of GOOSE and FOOT uttered by a gender-ambiguous voice, while being primed with an image of a male or female face. The results show no significant difference between the male face and female face conditions, although male participants identified significantly more fronted tokens as GOOSE and FOOT. In addition, participants accepted fronter tokens as GOOSE and FOOT when the face in the image was of the opposite gender to themselves. I contextualise my findings in terms of several socio-cultural and psychological factors that may have affected the variation in perception, focusing on the notion that people tend to associate linguistic change with 'other' people who are unlike themselves.

References

Fabricius, A. (2007). *Vowel formants and angle measurements in diachronic sociophonetic studies: FOOT-fronting in RP*. Paper presented at the International Congress of Phonetic Sciences, August 6-10, 2007, Saarbrücken.

Fridland, V. (2012). Rebel vowels: how vowel shift patterns are reshaping speech in the modern south. *Language and Linguistics Compass, 6*(3), 183-192.

Haddican, B., Foulkes, P., Hughes, V., & Richards, H. (2013). Interaction of social and linguistic constraints on two vowel changes in northern England. *Language Variation and Change, 25*(3), 371-403.

Harrington, J., Kleber, F., & Reubold, U. (2007). */u/-fronting in RP: a link between sound change and diminished perceptual compensation for coarticulation?* Paper presented at the International Congress of Phonetic Sciences, August 6-10, 2007, Saarbrücken.

Harrington, J., Kleber, F., & Reubold, U. (2008). Compensation for coarticulation, /u/-fronting, and sound change in standard southern British: an acoustic and perceptual study. *The Journal of the Acoustical Society of America, 123*(5), 2825-2835.

Kleber, F., Harrington, J., & Reubold, U. (2012). The relationship between the perception and production of coarticulation during a sound change in progress. *Language and Speech, 55*(3), 383-405.

Williams, A., & Kerswill, P. (1999). Dialect levelling: change and continuity in Milton Keynes, Reading and Hull. In P. Foulkes & G. J. Docherty (Eds.), *Urban voices: accent studies in the British Isles* (pp. 141-162). London: Arnold.

Uptalk in Standard Southern British English

Amalia Arvaniti & Madeleine Atkins
University of Kent
a.arvaniti@kent.ac.uk mga6@kentforlife.net

This study deals with the realization and function of uptalk in Standard Southern British English (SSBE). Uptalk has been extensively investigated in Australian, New Zealand and American English (among many, Fletcher, Grabe & Warren, 2005; Ritchart & Arvaniti, 2014), but, with the exception of Barry (2008), its use in the UK has been relatively understudied.

The present study examined uptalk in the speech of eight middle class speakers from South-East London and Kent. Four speakers (2F, 2M) were in their 20s, the other four (2F, 2M) in their 50s. They were recorded in their own homes while playing Cranium (a board game) and taking part in a Map Task. The aim was to determine whether speakers use uptalk and if so of what form and for what function(s). The data were analysed using the Autosegmental-Metrical framework for intonational phonology (Pierrehumbert, 1980); measurements of rise excursion, duration and velocity were also obtained.

Uptalk was used in 15% of the annotated utterances (N = 356). Uptalk frequency was not affected by age or gender, though female participants did use uptalk significantly more for floor holding (statements non-final in a turn). Uptalk was four times as frequent in the Map Task, where conversation is cooperative, than in the (competitive) board game. The majority of tokens were used for floor holds (33%) and statements (36%), with the remainder evenly divided between direct and indirect questions. Phonologically, two main phrasal melodies were present: L-H% used primarily for statements, and H-H% used primarily for direct and indirect questions. These melodies are illustrated in Fig. 1, taken from a Map Task session between the two older women: the H* L-H% melody indicates the instruction giver intends to continue, as indeed she does after a short pause. The H* H-H% melody is used twice in a row and indicates that the speaker is uncertain both about how to describe the image she sees and about whether her interlocutor has something similar on her map. The follower clearly interprets the last two instances of uptalk in this way, i.e. as indirect questions, searches for flowers on her map and responds with "roses" (where H* L-L% serves to indicate that she does have flowers, but not lillies). Finally, acoustic analysis of the rises indicates that indirect questions (like those on "the lillies" and "the flowers" in Fig. 1) had rises that were significantly longer in duration and had lower F0 velocity than rises used for other discourse functions.

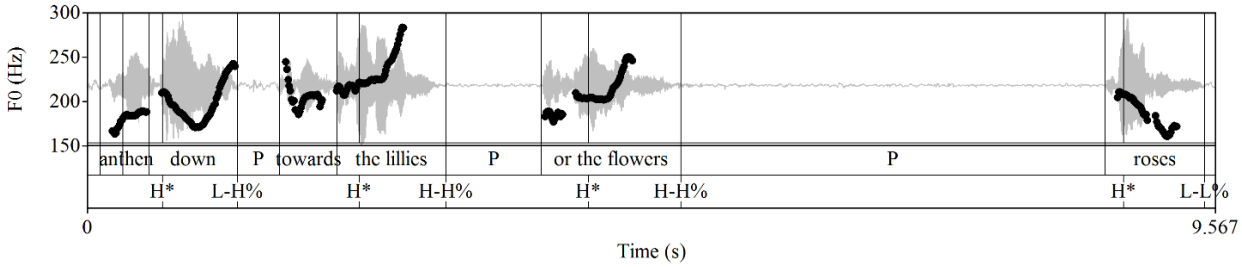


Figure 1: Waveform, pitch track and AM annotations illustrating the use of H* L-H% and H* H-H% in the data.

These results, though preliminary, indicate stable patterns of uptalk and can be juxtaposed to the variability reported by Barry (2008) for speakers of similar background. The phonological stability and lack of age effects could indicate that uptalk is becoming a stable SSBE feature, possibly by extending traditional uses of pitch rises in SSBE to new domains: H-H% has long been reported to be a means of expressing uncertainty and negotiating common ground in several varieties of English including SSBE, while the use of L-H% utterance-finally could be an extension

of the use of rising pitch utterance-medially (e.g. across an intermediate phrase boundary) to intonational phrase-final position to indicate the continuation of one's turn.

References

Barry, A. S. (2008). *The form, function, and distribution of high rising intonation: A comparative study of HRT in Southern Californian and Southern British English*. VDM Verlag Dr. Müller.

Fletcher, J., Grabe, E., & Warren, P. 2005.Intonational variation in four dialects of English: the High Rising Tune. In S. Jun (ed.), *Prosodic typology*, pp. 390-409. Oxford University Press.

Pierrehumbert, J. 1980. *The phonology and phonetics of English intonation*. Ph.D. thesis, MIT.

Ritchart, Amanda & A. Arvaniti. 2014. The form and use of uptalk in Southern California English. *Proceedings of Speech Prosody 7*, Dublin.

The acquisition of the pharyngealised fricative sounds amongst L2 learners of Arabic

1st Hajar Binasfour, PhD student, University of Reading
2nd Prof. Jane Setter, Professor of Phonetics, University of Reading

It has been argued that Arabic pharyngealised sounds are distinctive and unique amongst all world languages (Jongman, Herd, & Al-Masri, 2007; Newman, 2002) and difficult for second language (L2) learners of Arabic to acquire (Alsulaiman, Ali, Muhammad, Al Hindi, Alfakih, Obeidat, & Al-Kahtani, 2014). The present study examined the pharyngealised fricative sounds /sˤ/ and /ðˤ/ as they were pronounced by a group of L2 learners of Arabic. The purposes of this study are to describe the way L2 learners of Arabic produce these unfamiliar sounds in order to investigate the difficult aspects of acquisition and also to examine the effects of technology-based instructions in enhancing learners' production of these sounds.

Second language acquisition studies that investigated the relationship between speech perception and production concluded that enhancing learners' perception of sounds leads to better production (Huensch, & Tremblay, 2015). In the current study, an experiment was conducted to investigate the extent to which of two different methods of pronunciation instruction, traditional and technology-based, can contribute more to accurate production. It also adapted the method used in Olson (2014) to investigate the extent to which using speech analysis technology (Praat) can help in noticing the difference between pharyngealised and non-pharyngealised sounds.

The participants were 46 first language (L1) speakers of Mandarin, Urdu and Tagalog who were L2 learners of Arabic. Half of them received pronunciation instruction using traditional methods, and half received technology-based instruction. Data were collected from recordings of Arabic words containing the pharyngealised fricatives /sˤ/ and /ðˤ/ taken during pre- and post-test conditions.

The results of the study indicated that the pharyngealisation posed problems for L2 learners of Arabic the main issue being the incorrect production of pharyngealised sounds as non-pharyngealised sounds. In addition, an amount of L1 transfer was observed. However, both groups of learners performed significantly better after the pronunciation instruction. There was no statistical difference observed between L2 learners who received the traditional pronunciation teaching method and those who received pronunciation training using Praat.

References

Alsulaiman, M., Ali, Z., Muhammad, G., Al Hindi, A., Alfakih, T., Obeidat, H., & Al-Kahtani, S. (2014). Pronunciation errors of non-Arab learners of Arabic language. *In Computer,*

Communications, and Control Technology (I4CT), 2014 International Conference on (pp. 277-282). IEEE.

Huensch, A., & Tremblay, A. (2015). Effects of perceptual phonetic training on the perception and production of second language syllable structure. *Journal of Phonetics*, *52*, 105-120.

Jongman, A., Herd, W., & Al-Masri, M. (2007). Acoustic correlates of emphasis in Arabic. In *International Congress of Phonetics Sciences* (pp. 913-916).

Newman, D. (2002). The phonetic status of Arabic within the world's languages: the uniqueness of the lughat al-daad. *Antwerp Papers in Linguistics. 100*, 65-75.

Olson, D. J. (2014). Benefits of Visual Feedback on Segmental Production in the L2 Classroom. *Language, Learning & Technology*, *18*(3), 173.

The interplay between external sandhi and glottalisation in English-accented German and in English: an ultrasound imaging study in progress

Maria Paola Bissiri and James M. Scobbie
CASL Research Centre, Queen Margaret University, Edinburgh, Scotland, UK
{Mbissiri|Jscobbie}@qmu.ac.uk

We present an ongoing study which investigates the interplay between external sandhi – specifically word-final /r/ – and word-initial glottalisation in English-accented German compared to native English in relationship to accent and phrasing.

In Southern British English word-final /r/ is usually not articulated, but it can surface if the following word starts with a vowel [2]. In German, word-initial vowels are glottalised in the majority of cases [3], while in English word-initial glottalisation is less frequent and motivated by phrase boundaries and prominence [4]. Glottalisation is traditionally supposed to block sandhi, however both phenomena could co-occur [5]. The interplay between external sandhi and glottalisation has not been yet throughly investigated [5]. Regarding external sandhi in the non-native langue, few studies have been dedicated to this topic so far and with conflicting results [6].

In order to investigate /r/-sandhi, we employ ultrasound tongue imaging. We recorded Southern British English native speakers reading English and German sentences. The sentences contained two-word-sequences, in which the first word ended with /r/ and the second word started with a low vowel, e.g. *cider apples* and *jeder Abend* (*every evening* in German). The second word (*apples* or *Abend*) was either accented or deaccented, and between the two words in the sequence there was either a phrase boundary or not, thus obtaining four possible sentence types.

We hypothesise that glottalisation and external sandhi overlap is possible, and that external sandhi will be blocked by phrase boundaries, not by glottalisations. We expect that external sandhi will be most frequent in the deaccented, no-phrase boundary condition, and that glottalisation will be most frequent in the accented, phrase-boundary condition. Finally, regarding transfer to non-native productions we have two alternative hypotheses: a) English speakers will transfer their habits related to glottalisation and external sandhi from their native language to their non-native productions, or b) English speakers will separate words in their German speech, thus obeying to the "word integrity constraint" in the interlanguage suggested by [1], which "treats every word as a separate unit and prevents the articulatory synchronization of sounds belonging to different words" [1, p.19].

Preliminary results of the study will be presented at the conference.

[1] Cebrian, Juli. 2000. Transferability and productivity of L1 rules in Catalan-English interlanguage. *Studies in Second Language Acquisition* 22(1). 1-26.

[2] Cruttenden, A. and Gimson, A.C. 1994. *Gimson's pronunciation of English* (fifth edition), revised by Alan Cruttenden. London: Edward Arnold.

[3] Kohler, Klaus. 1994. Glottal stops and glottalization in German. Data and theory of connected speech processes. *Phonetica* 51. 38-51.

[4] Dilley, L., Shattuck-Hufnagel, S. and Ostendorf,M. 1996. Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics* 24(4). 423-444.

[5] Scobbie, J., Pouplier, M. (2010). The role of syllable structure in external sandhi: An EPG study of vocalisation and retraction in word-final English /l/. *Journal of Phonetics* 38, 240-59.

[6] Zsiga, E.C., 2011. External Sandhi in a Second Language: The Phonetics and Phonology of Obstruent Nasalization in Korean-Accented English. *Language* 87(2), 289-345.

Speech style as a predictor of /s/ variation across multiple contexts

Zac Boyd and Lauren Hall-Lew
The University of Edinburgh

The present paper analyses one talker's /s/ realisation across multiple different styles and two different points in time. Though a great deal of research has examined /s/ variation, little has focused on effects of speaking style. One exception is Maniwa *et al.* (2009), who found systematic differences in /s/ production between read and conversational speech. Here we contribute to the finding that /s/ is produced with a higher Centre of Gravity (CoG) in read tasks, but we also find differences between a variety of 'spontaneous speech' tasks.

The data come from one speaker from San Francisco, California and interviewed in 2012 and 2013, with a focus on stylistic variation in 2013 (Table 1). INTERVIEWS and READING PASSAGES are typical tasks in a sociolinguistic interview, but the 2013 reading passage data are unusual in that seven different passages were collected for comparison. The SELF-RECORDINGS encompass three recordings with four interlocutors made without the researcher present, a method used in recent sociolinguistic work (e.g., Podesva 2007). The LAB TASKS are nine tasks more or less typical of lab-based elicitation (e.g., a Map Task; Brown *et al*. 1984), wherein the speech elicited might be monologic or dialogic, and partly spontaneous and/or partly scripted. Previous analysis of this same talker found significant differences in her vowel production across the style contrasts from the 2013 data (Boyd *et al*. 2015).

	Interview	*Reading Passages*	*Self Recordings*	*Lab Tasks*
2012	✓	-	-	-
2013	✓	✓	✓	✓

Table 1: Time & style contrasts analysed with respect to /s/ production

All speech was transcribed and phone-aligned with FAVE (Rosenfelder *et al*. 2014) resulting in 2,449 /s/ tokens, and /s/ CoG data was extracted by Praat script. Peak frequency, kurtosis, skewness, duration, intensity and mean harmonicity were also measured but not analysed here. Regression models include word as random intercept. Though preceding and following phonological context were not included in the model at this time, all *str*- clusters were removed from the dataset (n=33) as these are known to retract.

The style analysis compares the four contexts from the 2013 data and finds significant differences between INTERVIEW speech and SELF-RECORDED speech, with the latter eliciting lower CoG values. We also find the expected differences between INTERVIEW speech and READING PASSAGES, with the latter eliciting higher CoG values. Interestingly, although the LAB TASK speech was expected to elicit a more spontaneous style than the READING PASSAGES, the /s/ CoG is essentially the same in both contexts.

Variation within each task type predicts /s/ variation as well, though the reasons are less clear. Comparing the 2012 and 2013 interviews, we find the speaker's /s/ CoG higher in 2013 than in 2012, which can likely be attributed to a more formal 2013 interview as part of a longer experiment, whereas 2012 occurred over lunch (i.e. less formal). There were no differences between the seven reading passages. In the self-recordings, the talker produced a more retracted /s/ with her sister than her friends. In the lab tasks we see a number of differences across task, but with no clear explanation (e.g., monologic tasks are not categorically different from dialogic tasks, nor are read vs. not-read tasks).

Overall, the phonetic realisation of /s/ varies widely but mostly predictably by broad speech style. While most work on /s/ CoG has focused on its indexical meanings for male talkers (c.f. Mack & Munson 2012), here we find significant intraspeaker variation for one heterosexual

woman. Some of our findings support previous results comparing read and non-read tasks. We also find robust differences for speech elicitation tasks not typically used in lab contexts, such as self-recordings (and tasks not typically used in sociophonetic fieldwork, such as the Map Task), and take this as a reminder to be cautious making conclusions about speech production on the basis of single-style speech.

Works Cited

Boyd, Zac, Elliott, Z., Fruehwald, J., Hall-Lew, L., & Lawrence, D. *An Evaluation of Sociolinguistic Elicitation Methods*. In The Scottish Consortium for ICPhS 2015 (Ed.), Proceedings of the 18th International Congress of Phonetic Sciences. Glasgow, UK: the University of Glasgow.

Brown, G., Anderson, A. H., Shillcock, R. and Yule, G. 1984. *Teaching Talk: Strategies for production and assessment*. Cambridge: Cambridge University Press.

Mack, S., & Munson, B. 2012. The association between /s/ quality and perceived sexual orientation of men's voices: implicit and explicit measures. *Journal of Phonetics,* 40*,* 198–212.

Maniwa, K., Jongman, A., & Wade, T. 2009. Acoustic characteristics of clearly spoken English fricatives. *The Journal of the Acoustical Society of America,* 125(6), 3962-3973.

Podesva, R. J. 2007. Phonation Type as a Stylistic Variable: The Use of Falsetto in Constructing a Persona. *Journal of Sociolinguistics*, 11, 478-504.

Rosenfelder, I., Fruehwald, J., Evanini, K., Seyfarth, S., Gorman, K., Prichard, H. and Yuan, J. 2014. *FAVE 1.1.3*. ZENODO. doi:10.5281/zenodo.9846.

L1 Tone attrition among bilinguals in an L2 speaking environment

Xiangjie Cao (Newcastle University)

Late bilinguals who continue to use their native language while using an L2 every day and/or residing in the L2 community have been shown to exhibit changes in their L1. The majority of the research on changes in L1 use and possible L1 attrition has focussed on the lexicon, morphology and syntax (Schmid 2002), but in recent years, attention has moved to phonology. Tonal attrition has received the least attention.

In Mandarin, tone is used to differentiate lexical items or to express morphological functions. There are four tones in Mandarin: the level first tone (T1), the rising second tone (T2), the falling-rising third tone (T3), and the falling fourth tone (T4). Among these, the tone considered to be the most complex is T3. Tone sandhi also applies to T3 where for two adjacent T3s, the first T3 is realized as T2 (Yip 1980). T2 and T4 show tone variations with different tones followed. In trisyllabic sequences, the middle T2 changes to T1 if the first syllable is T1 or T2 and the final syllable is a random tone from four tones.

Several studies over the past decade of Mandarin bilinguals have revealed attrition of tone by L1 Hakka Chinese speakers living in a Mandarin-speaking area (Yeh, 2011). Little is known, however, about what happens when a tone language speaker moves to a non-tone language environment. The present study addresses whether there are changes in tone production and perception by Mandarin speakers living in a non-tone language speaking environment (the UK) for varying lengths of time. The study compares 50 Mandarin-English late bilinguals who had been living in the UK from three months to more than five years with Mandarin monolinguals (only with minimal English exposure at school) living in mainland China. Their perception and production of four tones at word and sentence level were tested by a listening comprehension task, an interview task, and a story-telling task for both formal and more casual speech. A questionnaire collected data on speakers' use of and contact with both languages.

The data were analysed acoustically using Praat (version 5.4.22) speech analysis software (Boersma & Weenink 2015), and statistical measurement revealed that late bilinguals who had lived in the L2 environment for over five years showed signs of attrition on T3, tending to omit the raising part in production, and the first T3 in tone sandhi. The bilinguals' four tones showed a tendency to merge, rendering them less distinctive than the control group's tone production. Age of arrival, amount and type of L2 exposure and of L1 contact showed correlations with tone attrition. Moreover, some patterns mimic tone acquisition (Li and Thompson 1976; Lin 1985; Chang 2014) indicating that markedness plays a role in both acquisition and attrition.

References:

Boersma, Paul & Weenink, David (2015). Praat: doing phonetics by computer [Computer program]. Version [5.4.22], retrieved [8/10/2015] from http://www.praat.org/

Chang, Y (2014). First language Taiwanese Tonal Attrition: revisiting first language attrition hypotheses and their relevance. *Proceedings of the 37th Annual Penn Linguistics Conference*, vol.20. no. 6.

Li, C. N. & Thompson, S. A. (1976). The acquisition of tone in Mandarin-speaking children. *Journal of Child Language*, 4(02), 1977, 185-199.

Lin, William C.J. (1985). Teaching Mandarin tones to adult English speakers: analysis of difficulties with suggested remedies. *RELC Journal*, vol. 16, no. 2.

Schmid, M. S. (2002) *First language attrition, use and maintenance: the case of German Jews in Anglophone countries*. Amsterdam: John Benjamin.

Yeh, C.H. (2011). Language attrition and tonal change in Hakka. *Proceedings of the Psycholinguistic Representation of Tone Conference*, Hong Kong, 111-114
Yip, Moira. (1980). *The tonal phonology of Chinese*. Massachusetts Institute of Technology, Doctoral Dissertation.

Comprehension of English intonation by Chinese EFL learners

Yiling Chen, Ghada Khattab and Jalal Al-Tamimi (Newcastle University)

A growing body of research is being dedicated to exploiting the phonetic implementations of English intonation by ESL/EFL Chinese learners. The work often entails a comparison of the production and perception of phonetic correlates of focus/nuclear accents and nuclear tones by L2 learners and native speakers (e.g. Wang et al. 2011; Wang et al. 2010; Chen 2008). Some consistent findings in production reveal that: 1. Chinese learners are more likely to rely on F0 to realize accentuation and rarely reduce unstressed syllables' vowel duration, and 2. Chinese learners primarily resort to high-level (H*) instead of low-rise (L*H) to make the pitch accents stand out. These strategies have also been found in their performance on perception tasks (Wang et al. 2011). However, investigating differences at the phonetic level alone is by no means enough to understand learners' acquisition of intonation; the extent to which learners understand the meaning of intonation patterns is equally worthy of investigation, so as to better pinpoint the source of their incapability of proper phonetic perception and realization (Mennen and de Leeuw 2014). Unfortunately, so far little has been done in this respect. This paper is part of a PhD project which aims to examine whether and how Chinese EFL learners' overall ability to comprehend of English intonation could be improved after an intensive training course. Since the training is still ongoing, this paper only reports on the results from the pre-training test.

Using similar methods to Atoye (2005) and Cruz-Ferreira (1987), the stimuli in this study were 20 naturally produced sentences by two native linguists. Each sentence had two different versions, with meaning contrasted only via intonation (i.e. locations of accentuation, phrasing, or nuclear tone). Learners were placed in a sound-proof speech lab in xx University, sitting in front of a laptop and being instructed on how to carry out the tasks using DMDX. In each trial, listeners saw a target sentence and two meaning options of that sentence on the screen for 10 seconds. They then heard a version of the sentence and were forced to choose the corresponding meaning option of that version within 30s by pressing the labelled key "A" or "B". Each sentence was repeated three times, in which half of the sentences were randomly picked to repeat meaning option A twice and B once, while the other half repeated meaning option A once and B twice, resulting in 60 trials in total. Thirty-five Chinese students doing an MA in TESOL and ten native self-claimed RP speakers were recruited for this study.

Statistical analysis by One-way MANOVA showed that Chinese learners' overall correct rate of the comprehension task was significantly worse than that of native speakers ($F(3, 41) = 30.668$, $p < 0.001$; Wilks' $\Lambda = .308$; partial $\eta 2 = .692$). Follow-up univariate ANOVAs showed that both accentuation and phrasing-related intonational contrasts were much harder for Chinese learners to comprehend compared to tonal contrasts, as the former two were performed significantly poorer than native speakers (($F(1, 43) = 20.668$, $p < 0.001$; partial $\eta 2 = .325$ for accentuation; $F(1, 43) = 84.519$, $p < 0.001$; partial $\eta 2 = .663$ for phrasing), while the later one was not (($F(1, 43) = .211$, $p < 0.648$; partial $\eta 2 = .005$). These findings are insightful for our further experiment which will examine the production of these three intonational features, delving deeper into the extent to which learners' comprehension and their production ability are correlated, which has still been under-researched. The data from the post-training test will help us understand whether and how intonation can be acquired. It will also contribute to the advancement of teaching materials and pedagogical applications of intonation to this particular group of English learners.

References:

Atoye, R. O. (2005). Non-native perception and interpretation of English intonation. *Nordic Journal of African Studies*, *14*(1), 26-42.

Chen, H. (2008). On Chinese EFL learner's English Intonation Patterns. *Shanghai: Shanghai Foreign Language Education Press*.

Cruz-Ferreira, M. (1987). Non-native interpretive strategies for intonational meaning: An experimental study. *Sound pattens in second language acquisition*, 103-120.

Mennen, I., & de Leeuw, E. (2014). 'Beyond segments: Prosody in SLA'. *Studies in Second Language Acquisition*, *36*(02), 183-194.

Wang Xia, Aijun Li, and Xiaoli Ji. (2010) "Perception and Production of Prominence Distribution Patterns of Chinese EFL Learners." In *Speech Prosody-Fifth International Conference*.

Wang, Xia, Yuan Jia, and Aijun Li. (2011) "A comparative study on accentuation implementation of Chinese EFL learners vs. American native speakers." In *Speech Database and Assessments (Oriental COCOSDA),* pp. 62-67. IEEE.

Gender and real time change in the Scottish Vowel Length Rule in Glasgow

Florent Chevalier (University of Poitiers) and Jane Stuart-Smith (University of Glasgow)

In Scotland, the question of vowel quantity is extremely relevant, considering Scottish English has its own pattern of vowel duration, referred to as the Scottish Vowel Length Rule, formulated by Aitken (1981). This pattern differs from the one prevailing in most varieties of English, since vowels are not lengthened before all voiced consonants (this pattern being known as the voicing effect, or low-level-lengthening) but only when they are followed by a voiced fricative or a morpheme boundary. Aitken's Law was initially thought to apply to all vowels, albeit strong evidence for this pattern was only found for /i ʉ ai/ (McMahon 2000).

Studies on the realisation of the SVLR led in high Anglo-English contact situations, for instance in the border town of Berwick (Watt and Ingram 2000) or in Edinburgh (Hewlett et al 1999), revealed a tendency for young speakers to follow the voicing effect pattern rather than the SVLR. However, a study on male speakers in Glasgow, where there is substantially less contact with Anglo-English (see Scobbie et al 1999), concluded that the SVLR is resisting to any shift towards low-level-lengthening there (Rathcke and Stuart-Smith 2015). Furthermore, it pointed out that the SVLR is strongest for the speakers born earliest, but only when the vowel is in phrase final position.

Considering women are known to lead sound change, we wondered whether the phonological situation was similar for female speakers – or if it had already undergone greater weakening. To make it as comparable as possible to the male speakers-based study, we used the same corpus of real time Glaswegian English, the same years of recording (1970s and 2000s), and the same age of participants (young and middle-aged for both decades of recording), which altogether represents 12 female speakers falling into four groups. We extracted and checked all stressed realisations of /i/ and /ʉ/, except those followed by /r/ or likely to be reduced. Then, we labelled all usable tokens according to segmental and morphological environment, as well as prominence in the sentence and phrasal position, which we may expect to interact with vowel duration. We used Linear Mixed Effects modelling in R considering the effects on duration of fixed factors of SVLR, vowel quality, position in phrase, prominence of syllable, speaker group, number of segments/syllables, lexical frequency, and random factors for speaker and word.

Our results show the expected lengthening for vowels in final position as well as vowels carrying the main pitch accent. The SVLR is still actively used for /i/ and /ʉ/. However, it is clearly weakening, in both apparent time and real time, with middle-aged speakers recorded in the 1970s producing much longer vowels than young women recorded in the same decade and middle-aged women recorded in the 2000s. As Rathcke and Stuart-Smith (2015), this difference in length is particularly observed in phrase final position, which demonstrates the importance of prosodic factors in change in vowel duration. Overall, there is also no clear evidence for the implementation of voicing effect in Glasgow. Interestingly, the weakening of the SVLR shows a generally similar pattern for female and male speakers, with a suggestion that women from the 2000s are slightly more shortening their vowels than their male counterparts.

References

Aitken, A.J. (1981). The Scottish Vowel Length Rule. In Benskin, M. & Samuels, M.L.(eds.) *So meny People, Longages and Tonges: Philological Essays in Scots and Mediaeval English presented to Angus McIntosh*. Edinburgh: The Middle English Dialect Project, pp. 131-157.

Hewlett, N., Matthews, B., and Scobbie, J.M. (1999): Vowel duration in Scottish English speaking children. *Proceedings of the XVth ICPhS*, pp. 2157-60.

McMahon, A. (2000). *Lexical Phonology and the History of English*. Cambridge: Cambridge University Press, pp. 140-205.

Rathcke, T., and Stuart-Smith, J. (2015). On the tail of the Scottish Vowel Length Rule in Glasgow. *Language and Speech*, 2015, pp. 1-26.

Scobbie, J.M., Hewlett, N., & Turk, A. (1999). Standard English in Edinburgh and Glasgow: the Scottish vowel length rule revealed. In P. Foulkes and G.J. Docherty (eds.) *Urban Voices: Accent Studies in the British Isles*. London: Arnold, pp. 230-245.

Watt, D. and Ingham, C. (2000). Durational evidence of the Scottish Vowel Length Rule in Berwick
English. In: Nelson, D. & P. Foulkes (eds) *Leeds Working Papers in Linguistics 8*, pp. 205-228.

Modelling the changing rate and direction of historical and prehistoric sound changes

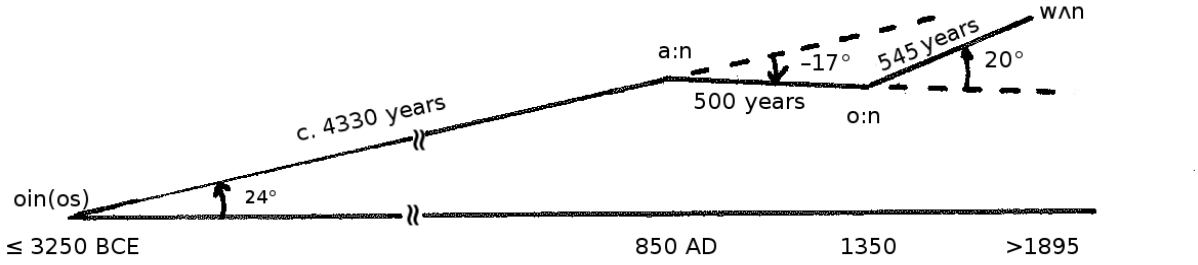
John Coleman (University of Cambridge)

The comparative method reconstructs hypothesised ancestral forms of words based upon their derived modern and/or historically-attested written forms. Developments in computational modelling of language phylogenies, inspired by computational models of biological evolution, enable hypotheses concerning ancestral forms, sound change rules, phylogenetic trees and chronologies to be tested (e.g. Nakhleh *et al*. 2005, Bouchard-Côté *et al*. 2013). In those works, the data used is text: alphabetic transcriptions of words and/or features represented linguistic traits. Recently, substantial progress has been made in phylogenetic modelling of how continuous functions (curves and surfaces) change and diverge across generations (e.g. The Functional Phylogenies Group 2012, Hadjipantelis 2013). As parameters of speech such as formant frequencies, amplitude contours, or surfaces such as spectrograms can be represented using continuous functions, it is becoming possible to model linguistic history and prehistory using these methods, and to reconstruct audible sound files instantiating hypothesized possible spoken forms from the past, including distant ancestral pronunciations and the intermediate forms at each generation.

As sound recordings from the distant past do not exist, we base our models on proxies of several kinds: (a) modern recordings that resemble how ancestral pronunciations are thought to have been; (b) hybrids, computed or (c) edited from two modern recordings; (d) statistical regression over a phylogenetic tree to extrapolate back from modern recordings to a hypothetical ancestral form. Our acoustic simulations of sound change are sometimes testable against observations or other sources of data. For example, the continuum from [ũ] (a supposed successor of Latin *un*-) to French [ɛ̃] includes intermediate stages like [œ̃], an attested, conservative French prounciation. However, linear interpolation between 'Latin' [tre:s] and French [tʁwa] does *not* pass through intermediate [trois], even though a form containing a diphthong [oi] and a final [s] is suggested by the historical spelling; the path of change is not linear in this case. A better understanding of the "landscape" of change is desired.

Many computational models of sound change assume that the rate of change is approximately the same everywhere and at all times. This is evidently not correct, as some words in some languages are more conservative whereas others are more innovative. For example, Lithuanian *penki* is much closer to Proto-Indo-European **penkwe* than is the English word *five*. Gradual, incremental modelling of sound changes in the acoustic domain enables us to estimate the varying rate at which words have changed over millennia. We characterise the direction of sound changes using cosines of spectral vectors; for example, the maximum spectral distance between Old English [a:n] and Middle English [o:n] is about –17° (minus because it is in the opposite direction from Proto-Indo-European **oin(os)* to [a:n]) over about 500 years. Such estimates depend somewhat on the choice of spectral representation, but tracing the twists and turns in the evolutionary pathways informs us about the shape of the landscape of change (Fig. 1). In this paper we present and discuss a selection of such temporal maps of changing pronunciations in the long history of English over timescales of hundreds and thousands of years.

Figure 1. Changing directions in the development of English "one" from Proto-Indo-European **oinos*



References

Bouchard-Côté, A., D. Hall, T. L. Griffiths and D. Klein (2013) Automated reconstruction of ancient languages using probabilistic models of sound change. *PNAS.* 110: 4224–4229.

Hadjipantelis, P.-Z. (2013) *Functional Data Analysis in Phonetics.* PhD thesis, Univ. of Warwick.

Nakhleh, L, D. Ringe and T. Warnow (2005) Perfect phylogenetic networks: a new methodology for reconstructing the evolutionary history of natural languages. *Language* 81 (2), 382–420.

The Functional Phylogenies Group (2012). Phylogenetic inference for function-valued traits: speech sound evolution. *Trends in Ecology and Evolution*, 27 (3), 160–166.

The phonetic and social correlates of non-rhoticity and derhoticised /r/ in Edinburgh English

Victoria Dickson[1] *and Lauren Hall-Lew*[2]
[1]*Faculty of Linguistics, Philology and Phonetics, University of Oxford, UK*
victoria.dickson@ling-phil.ox.ac.uk
[2] *School of Philosophy, Psychology and Language Sciences, University of Edinburgh, UK*
lauren.hall-lew@ed.ac.uk

The study of rhoticity has attracted a lot of interest in sociophonetic studies of urban Scottish speech. This work indicates a complex patterning in the realisation of postvocalic /r/ across the socioeconomic spectrum. The evidence points to rhotic approximants occurring in Middle Class speech, and derhoticisation (or a pharyngealized vowel; Speitel & Johnston 1983) typifying Working Class speech, both in Edinburgh (Romaine 1978; Lawson et al. 2008, et seq.; Scobbie et al. 2008, 2013; Schützler 2010, et seq.) and Glasgow (Macafee 1983; Stuart-Smith 1999, et seq.). Lawson et al. (2014: 63) model this Scottish /r/ variation along a 7-step phonetic continuum from deletion to derhoticisation to several approximant types and lastly to taps and trills. Recent research (Dickson & Hall-Lew 2015) has demonstrated that upwardly mobile speakers, who are intermediate in the socioeconomic scale, show the highest rate of approximant use, which suggests that rhoticity has a non-linear relationship with class. Furthermore, although acoustically similar, the non-rhotic and derhotic variants are maximally distinct in social distribution, used respectively by Middle Class women and Working Class men. Despite this, auditorily distinguishing non-rhotic and derhotic variants is notoriously difficult, such that even phonetically trained Edinburgh natives can struggle with coding them reliably (cf. Stuart-Smith et al. 2014). Thus, while Lawson et al. have evidenced a clear articulatory difference between the non-rhotic and derhotic forms, the question remains as to whether the variants do constitute auditorily distinct realisations of /r/, or whether instead the indexicality of a post-vocalic /r/ is signalled entirely by the quality of the preceding vowel.

We pursue this question by examining two of the several acoustic measures of derhotic /r/ described by Stuart-Smith et al. (2014) in a corpus of spontaneous speech from 16 male and female Edinburgh speakers aged 57-69, recorded in 2013. The sample represents three broad socioeconomic groups: the Working Class (WC), the upwardly mobile New Middle Class (NMC), and the Established Middle Class (EMC). Our first analysis consists of identifying instances of the 'breathy period' (Lawson et al. 2008) or 'audible frication' (Stuart-Smith et al. 2014) that is often found to occur for derhotic variants at the vowel offset, coding for its presence/absence and duration. In our second analysis we measured the F1 and F2 of the approximate midpoint of the preceding vowel for the subset of tokens belonging to the START lexical set.

We find that in 135 utterance-final tokens from 13 speakers, 51% of the data were realised with some frication. 87% of the fricative /r/s were produced by Working Class speakers and 90% were produced by men. The data also indicate that, overall, the duration of frication is longest for the WC speakers. While the presence or absence and duration of this breathy period seem to be a promising cue for distinguishing the variants, a number of complications arose with this analysis. Namely, the derhotic breathy period is really only discernible for utterance-final tokens and, among those, the duration measurements are highly variable, partly due to differences in amplitude quality as captured by the field recorder. The results for the START vowel show a strong interaction between gender and social class. The EMC women favour a higher, backer realisation of /ɑ/ (consistent with non-rhotic Anglo-English speech) compared to WC women, as was expected. The men on a whole produce a START vowel that is somewhere in between, and show little to no F1 or F2 differences according to socioeconomic class. Overall, we take these results to suggest that there are cases where it is possible to use acoustic measures to distinguish non-rhotic and derhotic variants, but in the (many) other cases where those measures are not feasible, the optimal definition of the /r/ variable considers the entire rime and takes the quality of the preceding vowel as a necessary parallel aspect of variation.

Selected References

Lawson, E., J. Stuart-Smith & J. M. Scobbie. 2008. Articulatory insights into language variation and change: Preliminary findings from an ultrasound study of derhoticization in Scottish English. *Selected Papers from NWAV 25, University of Pennsylvania Working Papers in Linguistics* 14: 102-109.

Lawson, E., J. M. Scobbie & J. Stuart-Smith. 2014. A socioarticulatory study of Scottish Rhoticity'. In R. Lawson (ed.), *Sociolinguistics in Scotland: Current Perspectives on Scots, Scottish English and Gaelic*. 53-78. Basingstoke: Palgrave Macmillan.

Macafee, C. 1983.*Varieties of English Around the World: Glasgow*. Amsterdam, The Netherlands: Benjamin.

Romaine, S. 1978. Postvocalic /r/ in Scottish English: sound change in progress? In P. Trudgill (ed.) *Sociolinguistic Patterns in British English*. 144-157. London: Arnold.

Schützler, O. 2010. Variable Scottish English consonants: the cases of /ʍ/ and non-prevocalic /r/. *Research in Language* 8: 5-21.

Scobbie, J., J. Stuart-Smith & E. Lawson. 2008. Looking variation and change in the mouth: Developing the sociolinguistic potential of Ultrasound Tongue Imaging: Full Research Report, *ESRC End of Award Report, RES-000-22-2032*. Swindon: ESRC.

Scobbie, J. M, E. Lawson & J. Stuart-Smith. 2013. *The rhotics and derhotics of Scottish English.* Lecture Handout, Yale University. New Haven, Connecticut.

Speitel, H. & P. Johnston 1983. *A sociolinguistic investigation of Edinburgh speech*. Unpublished Economic and Social Research Council (ESRC) end of grant report. Swindon, UK.

Stuart-Smith, J. 1999. Glasgow: Accent and Voice Quality. In P. Foulkes & G. Docherty (eds). *Urban Voices*. London: Arnold. 203-223.

Stuart-Smith, J., Lawson, E., and Scobbie, J.M. 2014. Derhoticisation in Scottish English: a sociophonetic journey. In: Celata, C. and Calamai, S. (eds.) *Advances in Sociophonetics*. Amsterdam: John Benjamins.

The effects of ambient language and age on the VOT of stop consonants in Bulgarian

Maria Dokovova, University of Cambridge

Voice onset time has long been considered one of the main cues for the voicing contrast in languages with a two- or three- way stop voicing distinction (Lisker and Abramson, 1964). English is usually described as an aspirating language, using mostly short-lag VOT for 'voiced' stops and long-lag (aspirated) VOT for 'voiceless' stops in word-initial position, while Bulgarian is considered a prevoicing language, using prevoicing and short-lag for the respective categories. It has been shown that speakers can experience changes in the VOT of their native language if exposed to a language, which realises voicing distinction differently (Sancier and Fowler, 1997). VOT realisation is also known to differ between age groups in English, with some evidence suggesting the possibility of a sound change (Docherty, et al. 2011; Stuart-Smith, et al. 2014).
The present study focuses on the VOT in the native language production of 10 Bulgarians living in Bulgaria and 10 in the UK and between two age groups within each country. The results show that the UK group produces voiced stops with shorter and less frequent prevoicing, than the Bulgarian group, the effect being stronger after longer residence in the UK. However, there are no significant differences in the production of voiceless stops between the two ambient language groups. Instead, there are significant differences in the VOT production between the two age-groups. The younger speakers tend to produce shorter prevoicing and longer aspiration than the older speakers, similar to what has been observed for English (Docherty, et al. 2011; Stuart-Smith, et al. 2014). The evidence of speech-rate differences between the two age groups is not consistent enough to account for the generational difference in the production. Diachronic data on Bulgarian stops is analysed to investigate the possibility of an ongoing sound change. Preliminary results support the previous evidence on the effects of the ambient language and a sound change in prevoicing. The results have important implications on the methodology of data elicitation and the phonetic features of Bulgarian.

References

Docherty, G., Watt, D., Llamas, C., Hall, D., & Nycz, J. (2011). Variation in voice onset time along the Scotish-English border. In *ICPhS XVII* (pp. 591–592).

Lisker, L., & Aabramson, A. (1964). A cross-language study of voicing in initial stops: acustical measurements. *Word*, *20*(3), 384–422.

Sancier, M. L., & Fowler, C. A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics*, *25*(4), 421–436. http://doi.org/10.1006/jpho.1997.0051

Stuart-Smith, J., Sonderegger, M., Mcdonald, R., Knowles, T., & Rathcke, T. (2014). The private life of stops VOT in real time corpus of Glasgowegian. In *The 14th Conference on Laboratory Phonology, Tokyo, July 25-27, 2014*.

Putting the larynx in the vowel space: Studying larynx state across vowel quality using MRI

John H. Esling[1], Scott Reid Moisik[2], Lise Crevier-Buchman[3,4]
esling@uvic.ca; Scott.Moisik@mpi.nl; lbuchman@numericable.fr

[1]University of Victoria; [2]MPI for Psycholinguistics, Nijmegen; [3]CNRS-UMR7018, LPP, Univ. Paris 3;
[4]Hôpital européen Georges-Pompidou

This study addresses how vowel quality interacts with larynx state. It is known that larynx height varies in relation to vowel quality, but inconsistently so (Ewan & Krones, 1974; Ladefoged, DeClerk, Lindau, & Papçun, 1972). Furthermore, some have suggested the intrinsic F0 of vowels arises from lingual-laryngeal interaction in producing vowels (Ohala, 1987; Whalen, Gick, Kumada, & Honda, 1998). Lingual-laryngeal interactions form an important part of Esling's (2005) Laryngeal Articulatory Model, which advocates a reshaping of vowel space organization to reflect the laryngeal component of vowel articulation, and Moisik (2013) proposes that lingual-laryngeal interaction emerges in numerous facets of phonological phenomena having to do with vowel quality in relation to laryngeal and pharyngeal consonants. Despite the evidence pointing to lingual-laryngeal interaction effects, uncertainty still remains about the nature of the articulatory changes involved, in particular those that concern the internal configuration of laryngeal tissues in response to vowel articulation.

Magnetic resonance imaging (MRI) lends itself well to studying lingual-laryngeal state because of its ability to capture rich information about the vocal tract in a safe way. We have thus acquired an MRI data set featuring key vowel qualities (produced by three trained phoneticians) in a number of different phonetic contexts captured using 2D (axial, coronal, and sagittal) multi-slice sequences and 2D midsagittal static and real-time sequences. The goal is to examine the variation in laryngeal structures as a function of vowel quality and the different contexts (including modal and creaky phonation, glottal and epiglottal stop, and raised larynx voice). Figure 1 provides an example of the data. In this talk we will provide an outline of this data set, details about our approach to its analysis, and some preliminary results concerning the quantification of larynx height and epilaryngeal cavity dimensions.

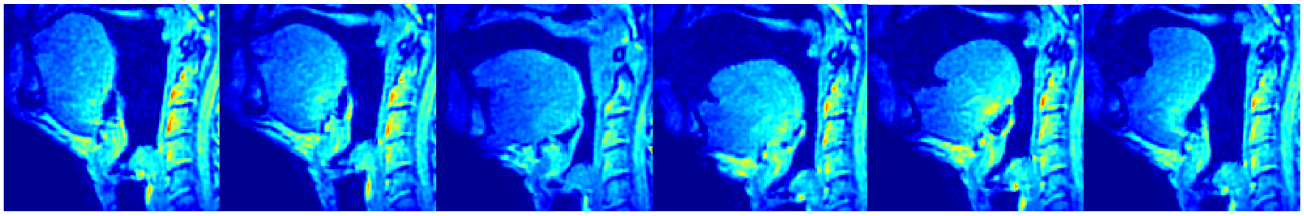


Figure 1: Midsagittal MR images of a phonetic modal vowel series produced with controlled phonatory quality and pitch (left of each image = anterior).

Esling, J. H. (2005). There are no back vowels: The laryngeal articulator model. *Canadian Journal of Linguistics*, *50*, 13–44.
Ewan, W. G., & Krones, R. (1974). Measuring larynx movement using the thyroumbrometer. *Journal of Phonetics*, *2*, 327–335.
Ladefoged, P., DeClerk, J., Lindau, M., & Papçun, G. (1972). An auditory-motor theory of speech production. *UCLA Working Papers in Phonetics*, *22*, 48–75.
Moisik, S. R. (2013). *The epilarynx in speech* (Doctoral dissertation). University of Victoria, Victoria, British Columbia, Canada.
Ohala, J. J. (1987). Explaining the intrinsic pitch of vowels. In R. Channon & L. Shockey (Eds.), *In honor of Ilse Lehiste* (pp. 207–215). Dordrecht: Foris.
Whalen, D. H., Gick, B., Kumada, M., & Honda, K. (1998). Cricothyroid activity in high and low vowels: Exploring the automaticity of intrinsic F0. *Journal of Phonetics*, *27*, 125–142.

Developing the vocal profile analysis scheme for forensic voice comparison

Peter French[1,2], Paul Foulkes[1,2], Philip Harrison[1,2], Vincent Hughes[1] and Eugenia San Segundo Fernandéz[1]

[1]J P French Associates, York, UK.
[2]Department of Language and Linguistic Science, University of York, UK.

Forensic voice comparison (FVC) involves analysing the speech of an unknown offender and a known suspect to aid the court in determining whether the voices belong to the same or different speakers. In the UK, FVC casework is conducted using a combination of auditory and acoustic linguistic-phonetic analysis. For forensic phoneticians voice quality (VQ) is considered one of the most valuable features for distinguishing between speakers (Nolan 2005). In a recent survey of practitioners, 94% report examining VQ, with 61% using a recognised framework such as Laver's VPA (Laver 1980).

However, for several reasons these frameworks are not widely used for full systematic VQ analysis. They require considerable training to use and can be relatively difficult to employ with typical forensic recordings (e.g. with poor technical quality, emotional speech, telephone transmission; Nolan 2005). The modified VPA used by J P French Associates in casework contains 38 separate dimensions, but some form a continuum (e.g. close jaw ~ open jaw) or are correlated (e.g. retracted tongue body ~ pharyngeal constriction). The inter-rater reliability of VPAs conducted using forensic material has not been reported, and the degree of within-speaker variability across multiple recordings has not been tested empirically. The present study addresses these issues and proposes developments for the VPA in FVC.

Data were taken from the 100 young male RP speakers in the DyViS corpus. High quality, near-end recordings of a telephone conversation from DyViS Task 2 were analysed. VQ was assessed by three analysts independently, using a modified version of the Laver (1980) VPA protocol. This had three rather than six scalar degrees for 'present' features, i.e. ignoring degrees which would be considered pathological. Features were scored on a 0 – 3 scale, where 0 means that a setting is absent and values of 1 – 3 mean the setting is present to an increasing degree.

Three sets of tests were performed. Firstly, correlations between each of the 38 VPA features were assessed. Empirical correlations were compared against predictions from phonetic theory to identify features which could be collapsed to simplify the scheme. Secondly, the inter-rater agreement across the three analysts was assessed. Initial results suggest a high degree of convergence across analysts, with relatively little disagreement about presence or absence of certain settings and typical differences of only one scalar degree for 'present' settings. Finally, a subset of 10 VPAs from Task 2 were compared with VPAs for the same speakers assessed from DyViS Task 1 (mock police interview) recordings in order to evaluate within- and between-speaker variation. Pairs of same- and different-speaker VPAs were reduced to a distance measure, and the distances used as speaker discrimination scores from which it is possible to generate error rates. These results represent a step towards formalising VQ analysis in FVC, and modifying the VPA scheme for wider forensic use.

References

Laver, J. (1980) *The Phonetic Description of Voice Quality*. Cambridge: CUP.

Nolan, F. (2005) Forensic speaker identification and the phonetic description of voice quality. In W.J. Hardcastle & J. Mackenzie Beck (eds.) *A Figure of Speech: A Festschrift for John Laver*. Mahwah NJ: Lawrence Erlbaum. pp. 385-411.

Dynamic Diachronic Formants

Josef Fruehwald (University of Edinburgh)

In this paper, I apply functional data analysis techniques (Ramsay & Silverman 2005) to the F1 formant tracks of /ay/ in order to investigate diachronic formant dynamics. In addition to the methodological benefits to be discussed, I have two empirical results to report from this exercise. First, pre-voiceless /ay/ underwent a change largely localised to its nucleus (F1 maximum), with marginal changes to its onset or offsets. Secondly, the effect of durational variation is very similar for pre-voiced and pre-voiceless /ay/ at the beginning of the vowel, but is more extremely different in the transition to the glide.

The preponderance of sociophonetic research on sound change characterizes vowels utilizing one time point representing the steady state's, or nucleus', formants and occasionally a second time point representing an offglide's formant targets. Formant dynamics beyond nucleus and glide estimates are being increasingly analyzed, typically involving extracting a fixed number of time points. Most commonly, five values at 20%, 35%, 50%, 65% and 80% are extracted, as in Jacewicz, Fox & Salmons (2011). With these additional time points, either averages are calculated at equivalent proportional time points, or non-linear smoothing is done, treating the formant estimates and associated time points as scalars. When examining how the formant dynamics are changing diachronically, an interaction with some diachronic measure may also be included.

In this paper, we treat formant estimates as "functional data", rather than as scalar data. That is, each observed formant estimate (formant_i), is a sample from a function of time $f(\text{time}_i)$. Each vowel token is converted from a scalar to a functional representation, then regression analysis can be used to examine how the functions have changed over time. This analytic approach has been utilized to analyze F0 contours (e.g. Gubian et al 2011), but not yet for diachronic vowel quality change.

In this analysis, I used 5,108 formant tracks of pre-voiceless /ay/ and 12,364 tokens of pre-voiced /ay/ extracted from the Philadelphia Neighborhood Corpus with a median number of 142 F1 estimates per token. These F1 values were z-score normalized based on the means and standard deviations from each speaker's entire vowel system. Each token was coverted to a functional representation using a 5th order b-spline basis with 15 knots, and a minimal roughness penalty determined by comparing Generalized Cross Validation estimates (Ramsay, Hooker & Graves 2009). This functional re-representation is conceptually similar to a Discrete Fourier Transform. The coefficients of these functions were then modelled in Stan (Stan Development Team 2014) with log2 vowel duration and speaker date of birth as the primary predictors, including random intercepts for speakers and words.

The results of the model are presented in Figures 1 and 2. In Figure 1, the effect of doubling duration on the F1 dynamics is presented for pre-voiceless /ay/, in blue, and pre-voiced /ay/, in red. As can be seen, the effect is nearly identical until about just after 25% of the vowels' durations, but a longer duration has a greater effect on the glide of pre-voiceless /ay/ than on pre-voiced /ay/. In figure 2, estimated F1 trajectories for pre-voiceless /ay/ are plotted for a subset of dates of birth in the left hand panel, and the difference from the estimated F1 trajectory for DOB=1900 in the right hand panel. The vowel quality change reached a maximum around DOB=1970, and as can be seen, is largely localised to the vowel nucleus, leaving the glide target largely unchanged.

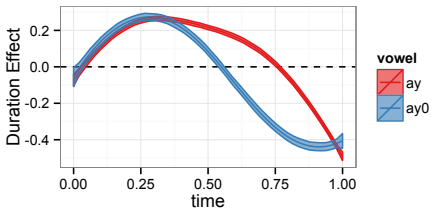


Figure 1: Dynamic duration effect.

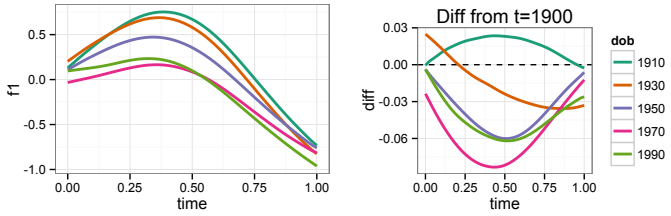


Figure 2: Dynamic F1 Change

Cheng, C. & Gubian, M., (2011)/ Predicting Taiwan Mandarin tone shapes from their duration. Interspeech 2011; Jacewicz, E., Fox, R. A., & Salmons, J. (2011). Cross-generational vowel change in American English. Language Variation and Change, 23(01), 45–86. doi:10.1017/S0954394510000219; Ramsay, J., Hooker, G., & Graves, S. (2009). Functional Data Analysis with R and MATLAB. New York, NY: Springer New York. doi:10.1007/978-0-387-98185-7; Ramsay, J., & Silverman, B. W. (2006). Functional Data Analysis. New York, NY: Springer New York.

The Categorical Perception of the /æ/-/ɛ/ continuum in British and North American English Speakers

Chad Hall (University of Oxford and University College London)

Abstract

In this study, the categorical perception of the /æ/-/ɛ/ continuum between British and North American English speakers was measured by participants performing a perception task. Because of the presence /æ/-tensing in many dialects of North American English, it was found that the North American speakers' boundary occurred later than British speakers due to their wider acoustic acceptance of /æ/. This is the first study to test the categorical perception of the /æ/-/ɛ/ continuum between British and North American speakers and the results are significant. The results also challenge the proposal that vowels are perceived continuously.

Affiliations: University of Oxford and University College London (UCL). This study was conducted at UCL in 2015.

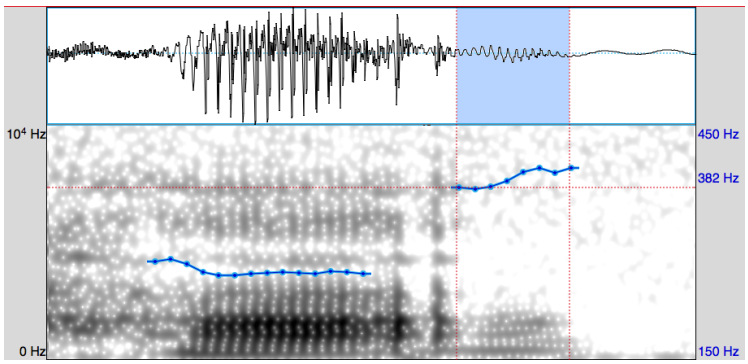
Acknowledgements: Thank you to Dr. Bronwen Evans, Mauricio Figueroa Candia, Gisela Tome Lourido, Sacha Babuta and Juan Pablo Ouviña Lanz for their valuable contributions to this project.

Glottal squeaks

Míša Hejná (University of Newcastle), Pertti Palo (Queen Margaret University), Scott Moisik (MPI for Psycholinguistics, Nijmegen)

This study sets out to explore whether the phenomenon referred to as "glottal squeak" (Anonymous 2015a & 2015b; Redi & Shattuck-Hufnagel 2001; Skarnitzl 2004) is conditioned by segmental, prosodic, and language-external factors in order to infer its articulatory characteristics.

Figure 1. Glottal squeak (taken from Anonymous 2015b).



Four aspects of glottal squeaks are considered: presence (i.e. frequency of occurrence), position with respect to sequences of vocalic and consonantal segments (e.g. at the juncture of a vowel and a consonant, in the middle of a vowel, etc.), duration, and f_0. Eight segmental conditions (vowel length, vowel backness, vowel height, vowel duration, F1, F2, manner of articulation of the consonant, place of articulation of the consonant) and three prosodic conditions are also taken into account (1. position within the foot: *pa*$^{\textbf{sqeak}}$*tter* vs *pa*$^{\textbf{squeak}}$*t*; 2. word uttered in isolation vs in the middle of a carrier sentence; 3. Stressed vs unstressed syllable: *pi*$^{\textbf{sqeak}}$*ck*, *pi*$^{\textbf{sqeak}}$*cker* vs *froli*$^{\textbf{sqeak}}$*c*). The analyses also focus on the presence of glottalisation, as defined by Redi & Shattuck-Hufnagel (2001), who found that squeaking is very rare, but "almost always" adjacent to other types of glottalisation (2001: 417). Furthermore, squeaking has been found to occur in female speech and to be speaker-specific. The data used here come from English spoken in Aberystwyth (10,006 tokens) and Manchester (410 tokens) and the analyses set out to provide further evidence on Redi & Shattuck-Hufnagel's findings.

The analyses of the presence of squeaks show that it is conditioned by the presence of glottalisation, sex, and manner of articulation of the consonant. Squeaks only occur *following* glottalisation in the plosive context and in female speech. They are also speaker specific. However, the low number of squeaks prevents any conclusive observations from being supported statistically.

Based on these results, squeaking may reflect the intrinsically tense vocal fold state associated with thyroarytenoid (TA) muscle recruitment (Esling, Zeroual & Crevier-Buchman 2007) required for epilaryngeal constriction and vocal-ventricular fold contact (VVFC) needed to produce glottalisation (Anonymous et al. 2015). In this interpretation, squeaks might occasionally occur during constriction disengagement: at the point when VVFC suddenly releases but the TAs have not yet fully relaxed. The gradually opening, but still narrow epilaryngeal tube may further contribute to the proneness to vibration (Titze 2008).

Variation in realizations of /r/ in Saudi Arabian Arabic: an auditory and acoustic study of two age-groups of female speakers.

Barry Heselwood, University of Leeds
Reem Maghrabi, King Abdulaziz University, KSA

The topic of this study is the realization of /r/ in Saudi Arabian Arabic as a function of prosodic context and speaker age, with a particular focus on the tap variant. Ten female native speakers were recorded reading aloud the Arabic version of the *North Wind and the Sun* passage (Thelwall & Sa'adeddin 1999: 53-4). Half of them were aged 22yrs, the other half 55yrs. The passage contains 14 instances of /r/ in six prosodic contexts: (1) foot-initial singleton after a consonant, (2) foot-initial intervocalic singleton, (3) foot-internal singleton after a consonant, (4) foot-internal intervocalic singleton, (5) foot-initial intervocalic geminate, and (6) foot-internal intervocalic geminate. Each subject read the passage twice. Discarding four reading errors, this yielded a total data set of 276 tokens of /r/. Through auditory analysis supplemented by inspection of waveforms and spectrograms, each token was classified into one of four general phonetic categories listed here as a hierarchy of increasing lenition: trill, tap, approximant, and zero realization. Subsequently, duration and intensity measurements were taken of tap realizations in order to investigate how much variation there can be in these acoustic parameters.

Results show the following overall percent occurrence of each category: trill (6.9), tap (65.2), approximant (23.2), zero (4.7). No overall differences in the frequency of the four classes across the two age-groups were found, but patterns of quasi-complementary distribution in the data set as a whole are evident, as is some difference between age-groups in some contexts. That is to say, certain realizations are more likely in certain contexts, and in one or other age-group, although not to the exclusion of other realizations. The percent occurrence of type by context is: (1) trill = 2.5, tap = 62.5, approximant = 25, zero = 10; (2) trill = 0, tap = 75.7, approximant = 18.9, zero = 5.4; (3) trill = 10, tap = 70, approximant = 15, zero = 5; (4) trill = 1.5, tap = 75.2, approximant, 19, zero = 4.4; (5) trill = 10, tap = 20, approximant = 70, zero = 0; (6) trill = 60, tap = 30, approximant = 10, zero = 0.

All realizations are subject to optional progressive devoicing after voiceless consonants. In some tokens, an epenthetic schwa-like vowel occurs between a voiceless consonant and /r/, particularly in the case of tap realizations, which has the effect of blocking devoicing.

Auditory and acoustic analysis led to the identification of subtypes of taps and approximants which themselves can be placed on a lenition hierarchy: affricated tap, 'strong' tap, 'normal' tap, weak tap (see Fig.1), 'tight' voiced approximant showing light frication, 'normal' approximant, weak or 'loose' approximant. Zero realizations can be seen as maximally weak approximants.

Regarding the extent of duration and intensity variation in taps, we find that durations vary between c.10–46ms, and intensity differences between voiced taps and adjacent vowels vary between c.3.0–16.0dB. We find that as intensity drop decreases, taps tend to sound weaker, suggesting a continuum in the amplitude of articulatory gestures for taps. The question of how auditory-perceptual categories map onto this articulatory continuum, and whether robust perceptual boundaries can be established for these subtypes, is identified as an issue for research into the articulation and perception of rhotic consonants.

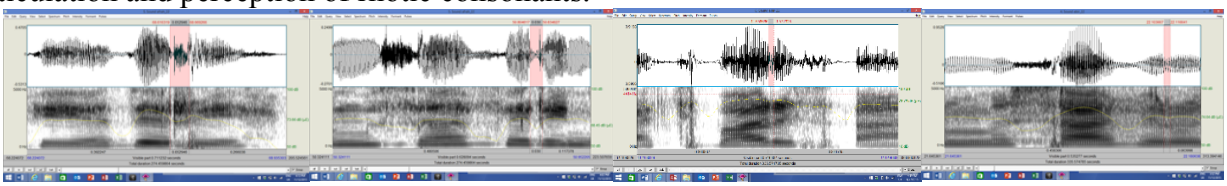


Fig.1. Highlighted tap realizations of /r/, from L to R: affricated, 'strong', 'normal', 'weak'.

Thelwall & Sa'adeddin (1999) Arabic. In *Handbook of the IPA*. Cambridge University Press. Pp.51-4.

Using ultrasound to look at the fluent speech of people who stutter

Cornelia J. Heyde[1], James M. Scobbie[1]

[1] *Clinical Audiology, Speech and Language (CASL) Research Centre, Queen Margaret University, Edinburgh, UK*

We present a new approach to analysing dynamic ultrasound data which we apply to the fluent speech of people who stutter (PWS) and control speakers (PNS).

Traditional approaches to analysing ultrasound involve the inspection of whole tongue surfaces at single points in time. Only recently have researchers moved away from the often impressionistic analysis of static data towards a more quantitative analysis of dynamic data. The tongue is a muscular hydrostat with an indefinite number of shapes and varying lengths. By its nature it therefore presents difficulties to researchers seeking to establish reference points at which measures should be taken. External reference points are not ideal, as they do not take into account articulatory and anatomical heterogeneity in a manner that permits inter-speaker/group comparisons.

The object of the current study is to present a reproducible approach to the analysis of dynamic ultrasound data [1] in order to test Wingate's Fault Line Hypothesis [2]. Wingate claims that disfluencies in the speech of PWS do not result from difficulty initiating speech, but instead when transitioning from syllable onset to rhyme. Disfluencies, according to Wingate, indicate difficulty integrating a syllable onset and the subsequent syllable rhyme.

We propose a method that uses the dynamic movement of the tongue to define the location of maximum movement. A measurement vector/referent is placed at this point. As the tongue moves along the measurement vector, movement strokes [3] of the tongue moving into (onset) and away from palatal constriction (offset) are registered. For each of the strokes we obtain measures for movement duration, distance and peak velocity.

We present the results for 9 PWS and 9 control speakers producing overall 321 repetitions of /ə kV/ with the vowel being /ɑ/ (n=112), /i/ (n=93) or /ə/ (n=116). As was hypothesised, results show comparable onset behaviours for both groups with none of the measures of lingual movement exhibiting significant differences. Regarding offsets, however, groups reveal significantly different movement patterns for stroke duration (ß=0.034 (se=0.012), t=2.75), average speed (ß=-12.765 (se=6.532), t=-1.954), and peak velocity (ß=-27.133 (se=11.804), t=-2.299).

These findings support Wingate's 'Fault Line Hypothesis' in that they suggest that PWS do not struggle initiating consonantal closure (onset). It is in the transition from consonantal closure into the vowel (offset), however, that PWS appear to employ a different strategy from PNS. The analysis technique developed in this study provides a novel and promising approach for exploring dynamic articulatory differences between speaker groups.

References

[1] Articulate Instruments Ltd 2012. Articulate Assistant Advanced User Guide: Version 2.14. Edinburgh, UK: Articulate Instruments Ltd.

[2] Wingate, M. E. (1988). The structure of stuttering: A psycholinguistic analysis. Springer Verlag, New York, NY.

[3] Tasko, S. M., & Westbury, J. R. (2002). Defining and measuring speech movement events. Journal of Speech, Language, and Hearing Research, 45(1), 127-142.

The phonetics and phonology of emphasis harmony in Qeltu and Gelet varieties of Iraqi Arabic

Maha Jasim, Newcastle University

Abstract

Emphasis harmony in vowels as driven by post-velar consonants has attracted the attention of many linguists working on Arabic (Zawaydeh, 1999; Shahin, 2003; among others) and other languages of Semitic (Rose, 1996) and non-Semitic origins (Wilson,2007). The production of post-velars involves a tongue body lowering or a tongue root retraction which imposes a featural change (lowering or retraction) in the neighbouring vowels in a process called emphasis harmony (cf. Ananian and Nevins, 2001). This research combines an experimental phonetic (acoustic) approach and a phonological feature-based approach to target emphasis harmony triggered by post-velars in the vowels of two Iraqi Arabic varieties, Qeltu and Gelet. The experimental approach involves embedding six target vowels /i a u/ and the their long counterparts /i: a: u: / in different word contexts; in each of the post-velar [ʕ], [ħ], [q], [χ], [ʁ] environments and in the environment of [tˤ] [ðˤ] [sˤ], the pharyngealised counterparts of [t] [ð[[s]. The word stimuli are introduced to twenty participants (10 from each variety with an age range 22-40) in a carrier sentence to elicit natural production of the target words in the speakers' own variety. The words are segmented and transcribed in Praat (Boersma and Weenink, 2015). The first three formants (F1-F3) are extracted at 25% and 50% of the vowel using a Praat script adopted and modified for the purpose of the study.

The acoustic findings are implemented in a feature model approach driven by previous approaches of Feature-Theory (Chomsky and Halle, 1968; Clements, 1985; McCarthy, 1988; Halle 1995). The research challenges the findings of previous theoretical accounts on emphasis harmony (Davis, 1995) showing that what has been stated do not match the acoustic findings in Qeltu and Gelet: the long /i (:)/ vowel shows some degree of lowering and retraction at the vowel onset in the post-velar environment (high F1 and low F2); thus it does not block emphasis harmony. However, the lowering and retraction of /i (:) / in the post-velar environment is more salient in F1 in Qeltu and in F2 in Gelet. Likewise, the /u (:) / is retracted in the post-velar environment in Gelet and lowered in Qeltu. However, the /a (:) / vowel shows higher degrees of retraction represented in vowel quality change from /a (:) / to [ɑ (:)] in the post-velar environment. Results show that this change of quality of the /a (:) / vowel is categorical in Gelet compared to Qeltu, and is not only restricted to the post-velar environment. This research concludes that all vowel environments in Qeltu and Gelet are transparent to emphasis harmony with an output of more retraction in Gelet and lowering of the vocals in Qeltu following the target post-velar consonants.

References

Ananian, C. S., & Nevins, A. I. (2001). Postvelar Harmonies: A typological odyssey. *Ms., MIT, Cambridge*.

Boersma, P. & Weenink. D. (2015). Praat. Doing phonetics by computer (v. 5.5.63). Retrieved 2 March 2015 from http:// www. Praat.org//.

Chomsky, N., & Halle, M. (1968). The sound pattern of English.

Clements, G. N. (1985). The geometry of phonological features. *Phonology*,*2*(01), 225-252

Davis, S. (1995). Emphasis spread in Arabic and grounded phonology. *Linguistic Inquiry*, 465-498.

Halle, M. (1995). Feature geometry and feature spreading. *Linguistic inquiry*, 1-46.

McCarthy, J. J. (1988). Feature geometry and dependency: A review.*Phonetica*, *45*(2-4), 84-108.

Rose, S. (1996). Variable laryngeals and vowel lowering. *Phonology*, *13*(01), 73-117.

Shahin, K. N. (2003). *Postvelar harmony* (Vol. 225). John Benjamins Publishing.

Wilson, I. (2007). The Effects of Post-Velar Consonants on Vowels in Nuu-chah-nulth: Auditory, Acoustic, and Articulatory Evidence. *The Canadian Journal of Linguistics/La revue canadienne de linguistique*, *52*(1), 43-70.

Zawaydeh, B. A. (1999). *The phonetics and phonology of gutturals in Arabic*. UMI Dissertation Service.

The iambic-trochaic law in Korean, Greek and English

Hae-Sung Jeon[1] & Amalia Arvaniti[2]
[1]University of Central Lancashire [2]University of Kent
[1]HJeon1@uclan.ac.uk [2]a.arvaniti@kent.ac.uk

The iambic-trochaic law (henceforth ITL) states that a louder sound signals the beginning of a group, while a longer sound signals its end. It is still not clear, however, whether ITL is a universal or the outcome of language-specific prosodic structures. Some studies have found no differences with respect to ITL between speakers of different languages; e.g. Hay & Diehl (2007) report no differences between English and French participants. Some studies, however, do report differences that appear to be based on the prosodic structure of the participants' language; e.g. in Iversen, Patel & Ohgushi, (2008) Japanese speakers showed no strong iambic preference with stimuli differing in duration as English speakers did. Results similar to those for Japanese are reported by Crowhurst & Olivares (2014) for Zapotec speakers.

The present study extends this line of research testing speakers of British English, Standard Greek and Standard Korean, languages that differ prosodically. British English uses intensity and duration to mark stress, has final lengthening and shows a preference for word-initial stress. Greek uses duration and intensity to a lesser extent than English to mark stress, has limited final lengthening and marks stress on one of the last three syllables of words. Korean, on the other hand, uses substantial lengthening to mark phrasal boundaries but does not have stress or any similar way of marking prominent syllables.

In the experiment, which ran on DMDX, Korean (N = 30), Greek (N = 25) and English (N = 28) listeners had to determine whether sequences of tones were grouped as trochees or iambs, choosing from visual representations of the two options (see Fig. 1). The stimuli varied to create three types of alternating sequences differing in: (i) duration; (ii) intensity; (iii) both duration and intensity so as to keep their summation constant (cf. Beckman, 1986). Generalised linear mixed-effect models were fitted to the data in R using lme4. The dependent variable was listener's choice (0: trochee, 1: iamb). Subject was a random factor; fixed factors included Sequence Type, Language, and Inter-Stimulus Interval (ISI). The best-fitting model was identified using log-likelihood χ^2 tests; *p*-values were corrected for multiple comparisons using the mcp function in the multComp package.

Fig. 1: Listeners' choice: (1) trochaic grouping, (2) iambic grouping

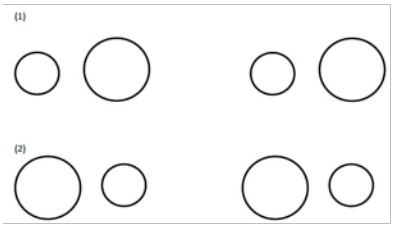
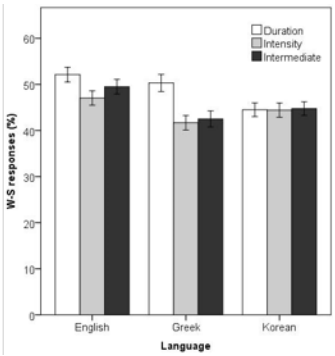


Fig. 2: Mean percentage of weak-strong (iambic) responses by language and stimulus type



The results revealed no significant cross-linguistics differences and only weak preferences in line with ITL (see Fig. 2). Importantly, grouping preferences were influenced by the duration of the Inter-Tone-Interval: when it was long (200 ms) participants were more likely to choose trochees than when it was short (20 ms). Considering previous studies and the present results together, the iambic-trochaic law seems to be a weak tendency, which may show cross-linguistic differences

depending on the precise nature of the stimuli, but do not support the view that the ITL is a general cognitive requirement.

References

Beckman, M. E. 1986. *Stress and Non-Stress Accent*. Dordrecht: Foris.

Crowhurst, M. J. & Olivares, A. T. 2014. Beyond the Iambic-Trochaic Law: The joint influence of duration and intensity on the perception of rhythmic speech. *Phonology* 31, 51-94

Iversen, J. R., Patel, A. D. & Ohgushi, K. 2008. Perception of rhythmic grouping depends on auditory experience. *Journal of the Acoustical Society of America* 124, 2263-2271.

Hay, J. F. & Diehl, R. L. 2007. Perception of rhythmic grouping: Testing the iambic/trochaic law. *Perception & Psychophysics* 69, 113-122.

Lenited vowels in Greek - experimental versus conversational speech

Marianna Kaimaki
University of Cambridge
mk724@cam.ac.uk

This paper examines the phonetic variability of Greek unstressed vowels in two speech modes: controlled experimental and unscripted conversational. Previous studies on the acoustics of Greek vowels have indicated that the high vowels /i/ and /u/ may have voiceless realisations in certain phonological contexts (Dauer 1980, Mennen and Okalidou 2006). However, more recent experimental work showed evidence that all Greek vowels can undergo lenition and become voiceless, providing that they are unstressed, preceded by a voiceless consonant, are sentence final and produced with falling pitch (Kaimaki 2015).

The current paper investigates whether the pattern of devoicing observed in the experimental data is also present in conversational speech. Unscripted conversations between friends and family members were searched for sentence (speaker turn) final words, in which the final vowels were preceded by either a voiceless plosive or voiceless fricative. The phonological criteria for including a word in the dataset were selected so as to match the experimental tokens used in Kaimaki 2015. Preliminary results from the analysis of 37 transcribed conversations confirm that all four Greek vowels, /i, ɛ, ɐ, ɔ/, may undergo lenition in conversational speech (there were no words with a final back vowel /u/ matching the criteria for selection in the transcribed portion of the data) but the range of possible phonetic realisations is different from the controlled experiment.

The vowels in the conversational data were predominantly voiced, which matches the pattern observed for the experimental data. Where there was lenition, however, the pattern did not match that found in the experimental data. Lenited vowels were realised in one of the following ways:

a) as voiceless vowels, during which the articulators maintain their position as for the corresponding vowel but there is no vocal fold vibration present,
b) as voiceless fricatives, during which the articulators do not change shape in order to produce a final vowel but instead maintain the stricture types for the voiceless fricatives preceding such vowels and
c) as absent vowels, during which there is no voicing, frication or other vocal activity observed after the consonant.

Results so far suggest that the mode of speech has an effect on the phonetic realisation of final Greek vowels, with experimental data restricting the variability in terms of lenition to voiceless vowels, while conversational speech favours more variability, with voiceless vowels, long fricatives and elided vowels being present in words produced in that mode.

References

Dauer, Rebecca. M. 1980. The reduction of unstressed high vowels in Modern Greek. *Journal of the International Phonetic Association* 10, 17-27.

Kaimaki, Marianna. 2015. Voiceless Greek vowels. In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: the University of Glasgow.

Mennen, Ineke & Areti Okalidou. 2006. Acquisition of Greek phonology: An overview. *Working Paper WP10*, Speech Science Research Centre.

Obstruent voicing, aspiration, and tone

James Kirby, University of Edinburgh <j.kirby@ed.ac.uk>

It has long been known that, in addition to salient differences in voice onset time (VOT), so-called 'voicing' differences between consonants are in fact signalled by a number of acoustic cues (Lisker, 1986). One cue that is often particularly salient in onset position is fundamental frequency (F0) at the onset of the following vowel, which is typically higher after voiceless than voiced sounds (House and Fairbanks, 1953 and much subsequent work). However, this *onset voicing effect* (OVE) does not seem to be universal, particularly in languages which contrast (phonetically) aspirated and unaspirated stops. For example, while F0 is reported to be higher after aspirated stops in languages like English, Korean, and Taiwanese, the opposite effect has been reported for Cantonese, Mandarin, and Bengali; whether or not a language is tonal may also play a role (see Hanson, 2009; Chen, 2011 for overviews).

Empirical clarity on this issue is important, because the presence (or absence) of the OVE can be used to infer information about the implementation of voicing both within and between languages. This paper presents a cross-linguistic comparison of the OVE in three languages: Khmer (Cambodian), an intonation language, and two tone languages, Central Thai and Northern Vietnamese. These three languages are of particular interest, because while all have a three-way contrast between (pre)voiced, voiceless unaspirated, and voiceless aspirated stops, Khmer is non-tonal, while Thai and Vietnamese are (genetically unrelated) tone languages. This design permits the use of an almost identical set of test items in all languages while 'varying' tonality, allowing us to (a) study the influence of aspiration on the OVE and (b) extend empirical coverage of the OVE in tone languages.

Methods. Audio recordings were made of 12-14 speakers of each language, producing 5 basic syllable types: the onsets /d t t^h n l/ + the low vowel /a:/. In Thai, these were crossed with 3 of 5 open syllable tones (high-falling, mid level, and low level), and in Northern Vietnamese with all 6 lexical tones, although here we restrict analysis to the three modally-voiced tones. Items were recorded three times each both (a) in isolation and (b) in an alternative-question carrier phrase appropriate for the language, with a fixed sonorant-final particle preceding the target. Both VOT and F0 measurements were taken at 10ms intervals beginning 40 ms before release of closure to 100 ms following release, the temporal range in which the OVE is most often observed (Hombert, 1978).

Results. In Khmer, F0 when following both /t/ and /t^h/ was ~15 Hz higher than when following /d n l/, persisting in both conditions at 50 ms. In addition, F0 was higher following /t^h/ than /t/, but only in isolation. In Thai and Vietnamese, no OVE was visible in the carrier phrase context at any time points; in isolation, F0 following /t^h/ was ~15 Hz higher than when following other onset types when occurring with a high-falling tone in Thai, while in Vietnamese this effect is only observed with the low-rising tone. In both languages, the effect is neutralized by 50 ms following voicing onset.

Discussion. While /t^h/ was generally found to raise F0, /t/ was found to pattern with /d n l/ in Vietnamese and Thai but with /t^h/ in Khmer. This suggests that it is not the presence or absence of aspiration itself, but rather the presence or absence of an active laryngeal spreading gesture (which may not always be accompanied by longer VOT) that accounts for the observed cross-linguistic variability. Similarly, while our data concur with previous findings that the OVE may be attenuated in tonal languages (Hombert, 1978; Francis et al., 2006), the fact that it is only visible on (isolated) high(-falling) tone in Thai but low(-rising) tone in Vietnamese suggests language-specific differences in the coordination of laryngeal gestures and their timing patterns, and warrants further investigation.

The Influence of the Mother Tongue on Rhythm Perception

Sumio Kobayashi
University of Kent, Canterbury, United Kingdom, sk532@kent.ac.uk

The purpose of this study is to explore the influence of the mother tongue on rhythm perception. In particular, the aim of the study is to examine the extent to which English and Japanese speakers are sensitive to differences between regular rhythm and irregular rhythm.

Rhythm irregularity can be divided into two types, clashes and lapses. An uninterrupted succession of accents, and the absence of accents create clashes and lapses respectively. In Japanese, unaccented (weak) syllables are frequent and irregularly distributed so that the rhythm becomes irregular. For instance, the rhythm of a phrase can be “W (weak) W W W W W W S (strong)” in reading a phrase “efferutou no ue” (top of Eiffel tower). However English rhythm tend to be regular, with alternations between stressed and unstressed syllables being mostly even. This is reflected in music as well. In Japanese traditional music, such as Gagaku, the rhythm can be 5/4 (2+3/4); in English (and all Western music of all types), regular rhythms such as 3/4 and 4/4/ prevail. Due to the familiarity with irregularities in both language and music, it is hypothesized that Japanese listeners would be more sensitive to differences between stimuli with regular and irregular rhythms than English listeners would (cf. Hannon & Trehub, 2005 and Patel, 2007).

My study investigates this hypothesis, using language, music and pure tones as stimuli, all with similar rhythmic structures. The aim of using three types of stimuli was to explore the connection between language, music and basic auditory processing. In experiments run through OpenSesame, Japanese and English native speakers listen to two sound files and rate the rhythmic difference between the first and second file on a scale from 1 to 6. Although in some trials, there is no difference between the two, the second sound file (test stimulus) in each trial is a variant of the first sound file (familiarization stimulus). The variant is made by removing one or two weak syllables from the familiarization file. Similarly syllable addition to the familiarization stimuli makes the second different. Familiarization stimuli are compared with all of the test stimuli within a group. For example, familiarization stimulus in isochrony (constant rhythm) group written below is compared with stimuli 1 to 4 of isochrony group.

Figure 1: Example of stimuli

Regular

```
Familiarization Stimulus
*    *    *    *
* *  * *  * *  * *
mama mama mama mama
H L  H L  H L  H L
Test Stimulus 1 is identical to familiarization stimulus.
Test Stimulus 2 (Preserved Metre)
*    *    *    *
* *  * *  * *  * *
ma:  mama mama mama
H    H L  H L  H L
Test Stimulus 3 (Lapse 1)
*    *      *    *
* *  * * *  * *  * *
mama mamama mama mama
H L  H L    H L  H L
Test Stimulus 4 (Lapse 2)
*    *      *    *
* *  * * *  * *  * * *
mama mamama mama mamama
H L  H L    H L  H L
```

Irregular

```
Familiarization Stimulus
*    *      *    *
* *  * * *  * *  * * *
mama mamama mama mamama
H L  H L    H L  H L
Test Stimulus 1 is identical to familiarization stimulus.
Test Stimulus 2
*    *      *    *
* *  * * *  * *  * * *
ma:  mamama mama mamama
H    H L    H L  H L
Test Stimulus 3 (Lapse 1)
*    *      *    *
* *  * * * *  * *  * * *
mama mamamama mama mamama
H L  H L      H L  H L
Test Stimulus 4 (Lapse 2)
*    *      *    *
* *  * * * *  * *  * * * *
mama mamamama mama mamamama
H L  H L      H L  H L
```

Higher ratings for test stimuli 3 and 4 than test stimuli 1 suggest the accuracy. Also higher rating for stimuli 4 than 3 reflects the accuracy. Pilot results show that Japanese listeners are more accurate than English listeners in detecting differences between regular and irregular rhythms in all

types of stimuli. This result indicates that the native language and its rhythmic characteristics affect rhythmic perception, regardless of the type of stimulus, and that there is no perceptual difference between lapses and clashes.

References

Hannon, Erin E., and Sandra E. Trehub. "Tuning in to musical rhythms: Infants learn more readily than adults." *Proceedings of the National Academy of Sciences of the United States of America* 102.35 (2005): 12639-12643.

Patel, Aniruddh D. *Music, language, and the brain*. Oxford university press, 2007.

The phonetics of negative responses

Emina Kurtić and Gareth Walker (University of Sheffield)

Negative responses to Yes/No-interrogatives (YNI) exhibit considerable phonetic variability. The potentially different conversational functions performed by this variability are not well understood (cf. recent work on phonetic variability and discourse function of "yeah": Freeman, Levow, Wright, & Ostendorf, 2015). A few findings are reported in small-scale studies that show e.g. that the phonetic characteristics of "no" indicate whether or not more talk will follow immediately from the same speaker (Ford, Fox, & Hellerman, 2004). Raymond (2010) argues that speakers also use prosody to register or project news, and to display a stance towards that news.

Our study is based on parametric phonetic analysis of all 133 instances of negative turn- initial "no" responses to YNI that could be found in the AMI Meeting corpus (Carletta, 2007). We take into account interactional organisation by considering whether talk from the same speaker immediately follows "no" (talk-projecting "no"), or not (turn-projecting "no"). We show that, on average, (i) talk-projecting "no" has a shorter duration than turn-projecting "no"; (ii) talk-projecting "no" has a higher mean intensity relative to the speaker's norm than turn-projecting "no"; (iii) talk-projecting "no" has a narrower F0 span than turn-projecting "no", and a higher mean F0 relative to the speaker's norm. We use decision tree classification to explore whether these durational, F0 and intensity features are relevant in discriminating between talk- and turn-projecting "no". The performance of the classifier and the resulting decision tree models offer insight into how phonetic design of "no" is relevant for turn or talk projection.

Carletta, J. (2007). Unleashing the killer corpus: Experiences in creating the multi-everything AMI Meeting Corpus. Language Resources and Evaluation, 41(2), 181–190.

Ford, C. E., Fox, B. A., & Hellerman, J. (2004). ''Getting past no": Sequence, action and sound production in the projection of no-initiated turns. In E. Couper-Kuhlen & C. E. Ford (Eds.), Sound patterns in interaction (pp. 233–269). Amsterdam: John Benjamins.

Freeman, V., Levow, G.-A., Wright, R., & Ostendorf, M. (2015). Investigating the role of 'yeah' in stance-dense conversation. In Proceedings of the 16th annual conference of the International Speech Communication Association (Interspeech 2015).

Raymond, G. (2010). Prosodic variation in responses: The case of type-conforming responses to yes/no interrogatives. In D. Barth-Weingarten, E. Reber, & M. Selting (Eds.), Prosody in interaction (pp. 109–129). Amsterdam: John Benjamins.

Talking in Time: a phonetically based training programme for cochlear implant users

Emina Kurtić and Bill Wells (University of Sheffield)

This paper describes Talking in Time (TinT) a computer based programme of self-administered rehabilitative conversational training for cochlear implant (CI) users. Taking a turn at the right time can be a challenge for CI users, especially if it necessitates talking in overlap with another speaker. Prosodic features, in particular F0 have been shown to be relevant to turn-end prediction (e.g. Boegels & Torreira, 2015) and realisation of turn-competition in overlap (Kurtic et al., 2013). However, despite significant progress in CI technology, F0 processing is not optimal, which is one of the reasons why CI users experience difficulty to fully participate in many everyday social encounters that involve multi-party conversations.
The development process of TinT was user-centered. 8 focus groups sessions with 5 adult CI users were conducted over a period of 1 year. In each session users engaged with and discussed a prototype of a turn taking task. Training sequences were drawn from naturally occurring conversations between four people, recorded as part of a cross-linguistic study of overlapping talk (Kurtic et al., 2012). Video recordings only capture the scene, but offer little opportunity for lip reading. This challenges the CI user to use phonetic cues and the available scene information for understanding turn-taking action.

Unlike most CI training software, which focuses solely on listening, TinT has both a listening and a speaking training component implemented in each of its 4 modules. Module 1 is an introductory module providing information on turn-taking and the structure of the training programme. Module 2 develops CI users' competence in "in-the-clear" situations where there is no conversational overlap. Users are asked to identify if turns are on time or late (listening) and to take a turn on time (speaking). Module 3 trains speakers for competitive turn incomings in overlap. Listening involves identifying phonetic (F0, intensity) and non-phonetic (semantics, pragmatic) cues to overlap competitiveness, and speaking requires taking a competitive turn. Finally, Module 4 trains users in non-competitive overlap that arises as a byproduct of conversation, then goes on to synthesise all turn-taking modes. The listening task asks users to differentiate between in-the-clear talk, accidental and competitive overlap. The speaking task simulates being involved in accidental overlap and asks the user to drop out upon becoming aware of the overlap. In each module users are offered immediate feedback on their performance and the opportunity to increase task difficulty by removing textual and visual information.

Bögels, Sara, and Francisco Torreira. 'Listeners Use Intonational Phrase Boundaries to Project Turn Ends in Spoken Interaction'. Journal of Phonetics 52 (2015): 46–57.

Kurtić, Emina, Bill Wells, Guy J. Brown, Timothy Kempton, and Ahmet Aker. 'A Corpus of Spontaneous Multi-Party Conversation in Bosnian Serbo-Croatian and British English', LREC 2012.

Kurtić, Emina, Guy J. Brown, and Bill Wells. 'Resources for Turn Competition in Overlapping Talk'. Speech Communication 55, no. 5 (June 2013): 721–43.

Motivations for variability in /r/-realisations: the case of Singapore English

Geraldine Kwek
Phonetics Laboratory, DTAL, University of Cambridge
gsck2@cam.ac.uk

Recent research on rhotics has focused on auditory and articulatory studies of the distribution and realisation of /r/ in UK varieties of English, showing variation determined by social class [2][6]. Others [4][5] have looked at acoustic and articulatory correlates of /r/ realisations in contact varieties in the UK (i.e. Bradford English spoken by Punjabi-English bilinguals) and propose variation in rhoticity to be representative of the speakers' acquiring of a contact variety rather than the effects of bilingualism [5]. The study of rhotics in Singapore's multiethnic and thus multilingual context provides a parallel environment of ethnic contact in which to study variation in /r/. Emergent rhotic features in Singapore English (SgE), a traditionally non-rhotic variety, include /r/-realisation in non-prevocalic positions [7] and, more recently, a labialized variant of the more common alveolar approximant [ɹ], the labiodental approximant [ʋ] [1]. It is hypothesized that the changes in rhotics of SgE will shadow those of the English varieties in the UK. This study thus focuses on an extensive auditory investigation of /r/-realisation in SgE, aiming to shed light on the factors motivating /r/ variation.

Audio data was collected from 40 male and female SgE speakers, aged 19 – 45, from Singapore's four major ethnic groups (i.e. Chinese, Malay, Indian, and Eurasian). Speakers were balanced for ethnicity, gender and age. All speakers had been educated in Singapore's bilingual education system in their foundation years, and read and speak English and at least one other language fluently. Most of them report English to be their dominant language. Audio recordings of read speech were made and tokens of /r/ in three phonological positions (i.e. word-/syllable-initial, intervocalic and consonant clusters) were elicited from each speaker. Auditory coding was then done on all tokens of /r/ in the stated contexts. Adapting scales used for English varieties in the UK [2][3][6], a 6-point index applicable to /r/-realisations in SgE was devised. All /r/ tokens were scored according to an auditory perception of the strength of their realisation: 5 – tap/trill [ɾ]/[r], 4 – strong alveolar approximant [ɹ], 3 – weak alveolar approximant, 2 – strong labiodental approximant [ʋ], 1 – weak labiodental approximant, 0 – null realisation (Ø).

Ethnically Chinese SgE speakers were found with the lowest average strength scores, followed by their Eurasian, Malay and Indian counterparts, respectively. It seems to suggest a process of /r/ lenition which is correlated with language background, with ethnically Chinese speakers leading the change. This could well be a natural occurrence of lenition encouraged by the patterns of language dominance and variation in rhoticity in the other languages spoken by SgE speakers. However, further analysis of individual sub-groups indicated an effect of phonological position where significantly more tokens of /r/ in consonant clusters had lower scores (i.e. Scores 0, 1 and 2). Additionally, the distribution of lower scores in female SgE speakers, particularly Score 2 (the labiodental approximant [ʋ]), suggests a spreading of weaker /r/ forms according to phonological positions; from consonant clusters, to word-/syllable-initial positions and finally to intervocalic positions. The fact that male SgE speakers also had many of their tokens of /r/ with lower scores in consonant clusters but few in word-/syllable-initial positions and none in intervocalic position corroborates with the hypothesis of this spreading of weak forms of /r/ conditioned by phonological positions. This study thus highlights the interactions amongst factors affecting /r/-variability, and the phonotactics of /r/ in SgE, setting the stage for further acoustic investigation which will

contribute to a deeper understanding of synchronic variation and diachronic sound change in rhotics in SgE.

References

[1] Deterding, D. 2007. *Singapore English.* Edinburgh: Edinburgh University Press Ltd.

[2] Dickson, V., Hall-Lew, L. 2015. Class, Gender and Rhoticity: The Social Stratification of Postvocalic /r/ in Edinburgh speech. *The 10th UK Language Variation and Change* (UKLVC), York.

[3] Foulkes, P., Docherty, G. J. 2000. Another chapter in the story of /r/: 'Labiodental' variants in British English. *Journal of Sociolinguistics* 4(1), 30-59.

[4] Hirson, A., Sohail, N. 2007. Variability of rhotics in Punjabi-English bilinguals. *Proceedings of the 16th International Congress of Phonetic Sciences* 1501–1504.

[5] Kirkham, S., Wormald, J. 2015. Acoustic and articulatory variation in British Asian English liquids. *Proceedings of the 18th International Congress of Phonetic Sciences,* Glasgow.

[6] Lawson, E., Scobbie, J.M., Stuart-Smith, J. 2011. The social stratification of tongue shape for postvocalic /r/ in Scottish English. *Journal of Sociolinguistics* 15(2): 256–268.

[7]Tan, Y.Y. 2011. To r or not to r: A sociophonetic analysis of /ɹ/ in Singapore English. *Proceedings of the 17th International Congress of Phonetic Sciences*, Hong Kong.

Obstruent effects on fundamental frequency: distinct aspects with distinct explanations

D. Robert Ladd and James P. Kirby, University of Edinburgh

It is well known (at least since House & Fairbanks 1953) that voice fundamental frequency (F0) is affected by supralaryngeal articulations, but the phenomenon is still remarkably poorly described and understood. This is partly due to divergent interests: phoneticians are primarily interested in the physical (biomechanical, aerodynamic, etc.) processes involved (e.g. Halle & Stevens 1971; Kohler 1982; Löfqvist *et al.* 1989); phonological theorists use evidence from such segmental effects to argue for specific featural analyses of consonant types (e.g. Kingston & Diehl 1994; Beckman et al. 2013); most intonation researchers treat such effects as noise in the F0 signal, to be controlled for in some way (e.g. Bruce 1977, Grønnum 1979); and most tone language researchers focus on the historical links between obstruent effects of F0 and tonal contrasts (but see Chen 2011). A paper on American English by Hanson (2009), though motivated primarily by phonetic interests, makes progress by introducing two types of controls: first, by investigating obstruent effects in carefully controlled intonational contexts, and second, by comparing the effect of obstruents against sonorants (which are assumed to have no effect on F0). Hanson shows clearly that F0 is locally raised following the release of a voiceless stop irrespective of aspiration (i.e. in both /CV/ and /sCV/ syllables), while F0 after a voiced (NB phonetically voiceless unaspirated) stop is comparable to F0 after a nasal.

This paper gives a report of our recent work on obstruent-related F0 perturbations. In the first phase of this study, we used Hanson's approach to investigate the F0 effects of voiced and voiceless obstruents in French and Italian, which, unlike English, contrast a fully voiced 'voiced' stop with an unaspirated (short-lag VOT) 'voiceless' stop. We found a pattern like what Hanson found in English, with F0 locally raised after voiceless obstruents and unaffected (relative to F0 after a nasal) after a voiced stop. However, because the voiced obstruents in French and Italian are fully voiced throughout the closure, we were also able to show that there is a substantial local *lowering* of F0 *during* the obstruent closure. These findings, summarised in Fig. 1, make two things clear: (1) 'voiceless' obstruents across languages (Fr. and It. vs. AmEng.) and across contexts (/CV/ vs. /sCV/ in AmEng.) show similar local raising of F0 irrespective of aspiration, which points to the effect of an active voicing-inhibition gesture (Hanson 2009, cf. Hoole & Bombien 2014); (2) the lowering effect of obstruent voicing is a distinct phenomenon with, presumably, a separate physical explanation (e.g., a side effect of laryngeal lowering to maintain voicing during closure: Bell-Berti 1975, Westbury 1983).

This second conclusion leads to the second phase of our study, currently in progress: if there are distinct explanations for F0 effects during and after obstruents, what happens to F0 *before* obstruents? Impressionistically, there is often an abrupt lowering immediately before an obstruent (Ladd 2008), but this has seldom been studied directly (though see Kohler 1982). Again using Hanson's approach of controlling intonational context and providing a sonorant benchmark, we investigate the effects of voicing, stress, syllable structure and intonational context on F0 in the vowel preceding an obstruent in English. Current (but still preliminary) results show small but clear effects of voicing, syllable structure, and prosodic context; *inter alia*, our data support the conclusion that voicing-induced F0 lowering is confined to the obstruent's closure interval, and suggest that Ladd's 'abrupt pre-obstruent lowering' is primarily an artefact of falling intonation.

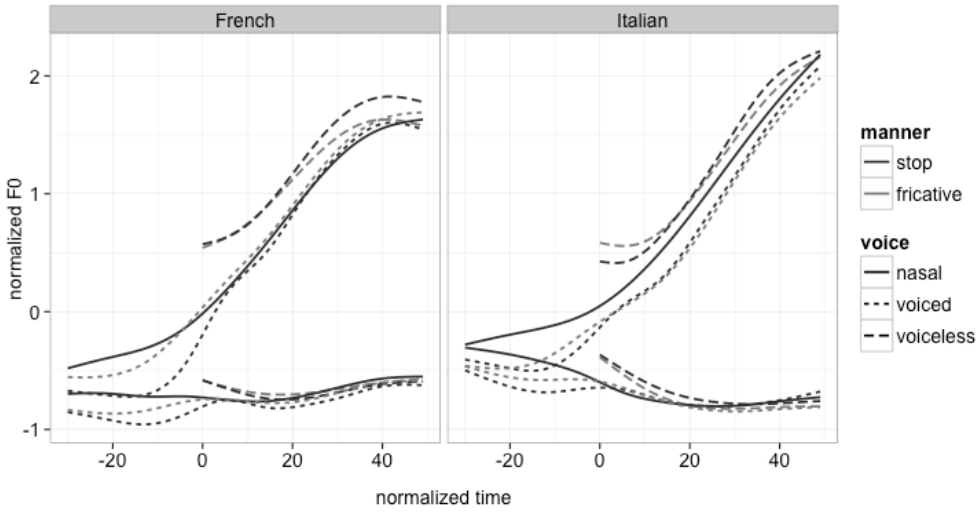


Fig. 1. F0 on word-initial CV sequences in rising (upper traces) and low falling (lower traces) intonational contexts.

References: Beckman, J. *et al.* (2013) *J.Ling.* 49:259-284.; Bell-Berti, F. (1975) *JASA* 57:456-461; Bruce, G. (1977) *Swedish Word Accents in Sentence Perspective* (Gleerup, Lund); Chen, Y. (2011) *J.Phon.* 39:612-625; Grønnum, N. (1979) *Phonetica* 36:57–78; Halle, M., & Stevens, K. N. (1971) *MIT Q. Prog. Rep*. 101:198–212; Hanson, H. M. (2009) *JASA* 125:425–441; Hoole, P. & Bombien, L. (2014) *Proc. 10th ISSP*, pp. 198-201; House, A. S., & Fairbanks, G. (1953). *JASA* 25:105– 113; Kingston, J., & Diehl, R. L. (1994). *Lg*. 70:419–454; Kohler, K. J. (1982) *Phonetica* 39:199–218; Ladd, D. R. (2008) *Intonational Phonology* (online appendix) (CUP: Cambridge); Löfqvist, A. *et al.* (1989) *JASA* 85:1314–1321; Westbury, J. (1983) *JASA* 73:1322-1336.

Why gesture delay leads to auditory /r/ lenition: an auditory, acoustic and articulatory study

Eleanor Lawson[1], Jane Stuart-Smith[2], James M. Scobbie[1]
[1]Queen Margaret University, Edinburgh; [2]University of Glasgow

The notion that there is a phonetic basis for the cross-linguistic tendency of coda consonants to lenite, vocalise or be deleted has, for a long time, been of interest to phoneticians. The possibility that a delayed anterior lingual gesture in coda liquid consonants might result in auditory weakening, or segmental loss, has been suggested by (Recasens and Farnetani 1994), who noted that the alveolar gesture of phrase-final /l/ in Catalan and American English was found to occur on occasion partially or completely after the offset of voicing, leading to segmental loss at the acoustic level, if not at the articulatory level. A similar phenomenon has been observed for Dutch /r/ (Scobbie et al. 2009; Scobbie and Sebregts 2011).

We investigate the contribution that anterior lingual gesture delay is making to the lenition of postvocalic /r/ in Scottish English. This study uses a socially-stratified, audio-ultrasound corpus of adolescent Glaswegian English (16 speakers), containing recordings from two sociolects; one with postvocalic /r/ weakening (working-class speech) and the other with /r/ strengthening (middle-class speech). We quantify auditory strength of rhoticity for all tokens of /r/ using a Praat-based rating experiment . Articulatory measures involved quantification of the temporal difference between the maximum of the anterior lingual /r/ gesture and either the offset of voicing in CVr words: *bar, bore, fur*, or the onset of a following labial consonant in CVrC words: *farm, herb, burp*. This articulatory measure we refer to as "lag". We considered both voicing offset and final consonant onset to be events that could auditorily mask the /r/ articulation to some extent. We also measured F1-F5, at, or as close as possible to, the /r/ maximum.

Correlation tests showed a strong correlation between a long, positive articulatory lag, a low /r/ index score (i.e. weak audible rhoticity) and a high F3. Linear mixed effects analysis shows that both social and linguistic factors significantly affect lag in our corpus. Working-class speakers have a positive lag that is significantly longer than middle-class speakers, who tend to have a short, negative lag. The presence of a pre-rhotic checked vowel /ʌ, ɐ, or ɜ˞/ results in a significantly shorter, negative lag for /r/, than the presence of a pre-rhotic nonchecked vowel. There was also a significant interaction between speaker social class and the presence of prerhotic checked vowels in the stimuli. This latter result is likely due to the presence of covert, bunched articulatory /r/ variants in the speech of the middle-class participants, as /r/-bunching has previously been shown to have a strong coarticulatory effect on preceding /ʌ, ɐ, ɜ˞/, resulting in vowel-consonant coalescence to [ɻ͡ʊ] (Lawson et al. 2013).

REFERENCES

Lawson, E., Scobbie, J.M. and Stuart-Smith, J. 2013. Bunched /r/ promotes vowel merger to schwar: An ultrasound tongue imaging study of Scottish sociophonetic variation. *Journal of Phonetics,* 41 (3–4) 0, pp.198-210.

Recasens, D. and Farnetani, E. eds. 1994. *Spatiotemporal properties of different allophones of /l/: phonological implications: Phonologica 1992: Proceedings of the 7th International Phonology Meeting.* Torino: Rosenberg & Sellier.

Scobbie, J.M. and Sebregts, K. 2011. Acoustic, articulatory and phonological perspectives on rhoticity and /r/ in Dutch. In: Folli, R. and Ulbrich, C. eds. *Interfaces in linguistics: New Research Perspectives.* Working Paper ed. CASL Research Centre Working Paper WP18, pp. 257-277.

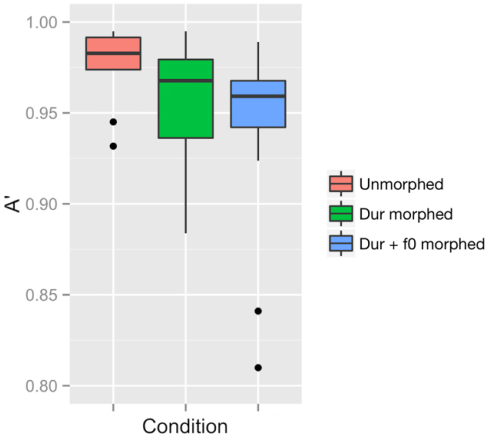
Scobbie, J.M., Sebregts, K. and Stuart-Smith, J. 2009. *Dutch rhotic allophony, coda weakening, and the phonetics-phonology interface.* working paper - 18 ed. Edinburgh: Queen Margaret University.

Disentangling the contribution of segments and prosody in dialect identification

Adrian Leemann, Francis Nolan (University of Cambridge)

The speech stream consists of segmental and suprasegmental features. Both of these levels carry diagnostic information for dialect identification. What current research has neglected to resolve is the question of the individual contribution of these cues in the identification process. We address this gap by means of a series of experiments in which segmental and suprasegmental cues are separated from one another. This can be achieved by manipulating the speech signal: the suprasegmental features of dialect X are morphed onto the segments of dialect Y and vice versa [cf. 1]. These morphed stimuli are presented to listeners who judge whether the sentence heard was from a speaker of dialect X or Y.

In a between-subjects design with three conditions (unmorphed, duration morphed, duration + f0 morphed), we recorded 6 Bern (BE) and 6 Valais (VS) Swiss German speakers. Previous research has reported differences in intonation [2], vocalic and syllable durations [3], and speech rate [2] for the two dialects. Each speaker read the same 10 sentences. The material was manually labeled in Praat [4]. For the morphing conditions, the respective suprasegmental features of sentence 01 of speaker BE01, for example, were morphed onto the segments of sentence 01 of speaker VS01; likewise, the suprasegmental features of sentence 01 of speaker VS01 were morphed onto the segments of sentence 01 of speaker BE01 etc. [5]. 60 Zurich German listeners (20/condition) participated in the experiment. Listeners heard 120 stimuli. Following stimulus presentation, subjects indicated whether the sentence heard was BE or VS Swiss German by clicking on the corresponding button on a laptop screen.



Results indicate that (i) listeners are highly sensitive in discriminating between the two dialects and (ii) suprasegmental cues seem to occupy only a marginal role when it comes to identifying a speaker's dialect. Figure 1 shows the boxplots of the dialects' A' [6] – the higher this measure, the higher identification performance (chance=.05.; 1=perfect identification). Listeners were able to identify the dialects much above chance in every condition. Yet, accent identification significantly deteriorated when syllable durations were swapped (green box, LM p=.03*) and, even more so, when syllable durations and f0 were swapped (blue box, LM p=.02*).

Figure 1: *Boxplots of A'.*

These findings help us better understand the role of segmental and suprasegmental cues in the identification of a speaker's dialect. A speaker's dialect significantly contributes to the uniqueness of a speaker's voice, which is why dialectal information is typically incorporated in the analysis of criminal and suspect recordings [7, 8]. The current study's results contribute to the increasing knowledge base on dialect identification and may enhance the diagnostic power of expert listeners' and earwitness's claims about suspect speakers' dialects.

References

[1] J. Vaissière and P. Boula de Mareüil, "Identifying a language or an accent: from segments to prosody", Paper read at *MIDL*, Paris 29.-30.11.2004.

[2]A. Leemann, *Swiss German Intonation Patterns*. Amsterdam / New York: Benjamins, 2012.

[3] A. Leemann and B. Siebenhaar, "Statistical Modeling of F0 and Timing of Swiss German Dialects", *Proceedings of Speech Prosody 2010.*

[4] P. Boersma and D. Weenink, *Praat: Doing phonetics by computer*. http://www.praat.org/, 2015.

[5] P. Boula de Mareüil and B. Vieru-Dimulescu, "The contribution of prosody to the perception of foreign accent", *Phonetica*, vol. 63, pp. 247–267, 2006.

[6] D. M. Green and J. A. Swets, *Signal detection theory and psychophysics*. New York: Wiley, 1966.

[7] M. Jessen, "Speaker Classification in Forensic Phonetics and Acoustics," in: C. Müller (Ed.), *Speaker Classification*, vol. 1, pp. 180-204, 2007.

[8] O. Köster, R. Kehrein, K. Masthoff, and Y. H. Boubaker, "The tell-tale accent: identification of regionally marked speech in German telephone conversations by forensic phoneticians*," Journal of Speech, Language and the Law*, vol. 19, no. 1, pp. 51–71, 2012.

Short term learning of derhoticised /r/ in Glasgow: Adaptation occurs for experienced and inexperienced listeners

Robert Lennon, Rachel Smith, Jane Stuart-Smith
Glasgow University Laboratory of Phonetics

In Glasgow, speakers are stereotypically rhotic. However, recent sociophonetic research indicates a trend towards the loss of postvocalic /r/ in working class Glaswegian speech, leading to 'derhoticisation' (Stuart-Smith 2007), alongside a simultaneous strengthening in rhoticity in middle class Glaswegian (Lawson et al. 2011; Lennon et al. 2015). The potential for misperception exists when listeners hear minimal pairs such as e.g. *cut/curt*, when spoken by working class speakers who realise postvocalic /r/ as a pharyngealized variant. The pharyngealized nature of derhoticised /r/ makes it perceptually very similar to the preceding open back vowel in /CʌrC/, leading to difficulty when listeners try to distinguish between /CʌC/ and /CʌrC/ words (Lennon 2014). Indeed, minimal pairs such as *hut/hurt* are also acoustically very similar in the Glaswegian working class accent (Lennon et al. 2015). Comparing native Glaswegian listeners, English listeners resident in Glasgow, and English listeners resident in Cambridge, Lennon (2014) showed that this difficulty decreases as experience of the Glaswegian linguistic environment increases. Furthermore, short-term exposure to unfamiliar phonetic detail can cause perceptual adaptation (e.g. Barden & Hawkins 2013; Eisner et al. 2013; Smith et al. 2014; Maye et al. 2008; Norris et al. 2003), therefore the present experiment tested the hypothesis that misperception of derhoticised variants of /r/ would decrease after a short period of exposure.

A three part experiment was designed:

1. Pretest: Listeners performed a 2-alternative-forced-choice (2AFC) task, with 96 stimuli: 12 target pairs (e.g. *cut/curt*) and 36 distractor pairs (e.g. *seem/same*, *bad/pad*) produced by a native working class Glaswegian male;
2. Exposure to read passage: Listeners heard a six-minute story read by the same speaker, which contained 12 pairs of /Cʌ(r)C/ words, (different words from those heard in Pretest and Posttest: e.g. *shut/shirt*);
3. Posttest: Listeners performed a second 2AFC (same design and stimuli), to compare responses from Pretest.

Two different versions of the Exposure story were created. In the Natural condition, the /Cʌ(r)C/ words were unmanipulated, while in the Altered condition they had the /ʌ(r)/ portion altered across three acoustic dimensions, in order to neutralise the differences between minimal pairs. Using Praat's source-filter synthesis to manipulate the real recordings, differences between F2, F3 and duration of all target word pairs were neutralised, the words were then recreated, and finally they were carefully inserted back into the passage so any transitions would be inaudible to the listeners. Thus groups hearing the Natural version of the story had the opportunity to learn the subtle phonetic cues distinguishing /Cʌ(r)C/ pairs, while those hearing the Altered version did not.

128 participants were from six listener groups, with varying levels of experience of Glaswegian. Each familiarity level had two groups - one heard the Natural story, the other heard the Altered version:

- 2 x Glasgow (raised in Greater Glasgow) n=(2x21) 42;
- 2 x Intermediate (raised in England, living in Glasgow, mean = 3.1 years) n=(2x22) 44;
- 2 x Cambridge (raised in South East England, very little experience of Glaswegian) n=(2x21) 42.

Results from Signal Detection Analysis (which measures sensitivity to differences (d') and response bias (c)) show that the Cambridge listeners were the least sensitive to the differences between e.g. *cut/curt*, Glaswegian listeners were the most sensitive, and Intermediate listeners patterned between the other groups, replicating the long-term learning pattern found by Lennon (2014). The short-term

learning results suggest that exposure to the Natural passage aids sensitivity to differences between minimal pairs for all listener groups, suggesting that listeners can improve their discrimination between fine grained phonetic features such as pharyngealized /r/ after only a small amount of exposure. The results for response bias are more complex. All groups, even the experienced (Glaswegian) listeners, change their bias after hearing the Altered story, suggesting they are learning that their normal cues are not relevant. These results appear to support the hypothesis that short-term learning, even with the limited exposure experienced by the listeners in this study, can promote a perceptual shift, which possibly indicates the early stages of ongoing learning.

References:

Barden, K., & Hawkins, S. (2013). Perceptual learning of phonetic information that indicates morphological structure. *Phonetica*, 70(4), 323-342.

Eisner, F., Melinger, A., & Weber, A. (2013). Constraints on the transfer of perceptual learning in accented speech. *Frontiers in psychology*, 4.

Lawson, E., Scobbie, J.M., & Stuart-Smith, J. (2011). The social stratification of tongue shape for postvocalic /r/ in Scottish English. *Journal of Sociolinguistics*, 15(2), 256-268.

Lennon, R. (2014). Increased exposure can aid perception of ambiguous /r/ variants in Glasgow. Paper presented at BAAP, Oxford University.

Lennon, R., Smith, R., & Stuart-Smith, J. (2015). An acoustic investigation of postvocalic /r/ variants in two sociolects of Glaswegian. *18th ICPhS*, Glasgow, Scotland.

Maye, J., Aslin, R. N., & Tanenhaus, M. K. (2008). The weckud wetch of the wast: Lexical adaptation to a novel accent. *Cognitive Science*, 32(3), 543-562.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive psychology*, 47(2), 204-238.

Smith, R., Holmes-Elliott, S., Pettinato, M., & Knight, R. A. (2014). Cross-accent intelligibility of speech in noise: Long-term familiarity and short-term familiarization. *The Quarterly Journal of Experimental Psychology*, 67(3), 590-608.

Stuart-Smith, J. (2007). A sociophonetic investigation of postvocalic /r/ in Glaswegian adolescents. *16th ICPhS*, Saarbrücken, Germany.

Aspiration effect on F0: language- and gender-specificity

Yang Li (University of Cambridge)

The effect of aspiration on fundamental frequency has been investigated in both tonal (Xu & Xu, 2003; Zee, 1980) and non-tonal languages (Hombert, 1975; Jeel, 1975). There is, however, a lack of consensus on the direction and scale of aspiration-induced F0 perturbation, primarily due to conflicting empirical findings, often within the same language (Lai, Huff, Sereno, & Jongman, 2009). For instance, aspirated obstruents in Beijing Mandarin and Shanghai Wu (Chen, 2011; Xu & Xu, 2003) seem to result in F0 onsets that are lower than their unaspirated counterparts. Data on Taiwanese Southern Min reveal a more complicated pattern: On average, initial aspirated obstruents have higher F0, in contrast to Mandarin and Shanghai; female Taiwanese speakers are more consistent in this effect than male speakers, who show significant inter-speaker variation.

The present study presents experimental data on 12 (6F, 6M; mean age = 39) simultaneous bilingual speakers of Fuzhou Min and Mandarin Chinese. Both languages feature an aspiration contrast in stops and affricates, and both have a high level tone and a high falling tone. The experiments took place over two sessions, where aspiration effect on F0 in each language is examined separately. In one session, participants were instructed to read two sets of eight monosyllabic words in Fuzhou, which differ minimally in the presence of aspiration in initial obstruents. In the second session, participants read similarly designed stimuli, except in Mandarin. Altogether, 1152 tokens were analysed from 12 speakers, 2 tonal conditions (high level & high falling), 2 languages, 8 test words (16 fillers), and 3 repetitions.

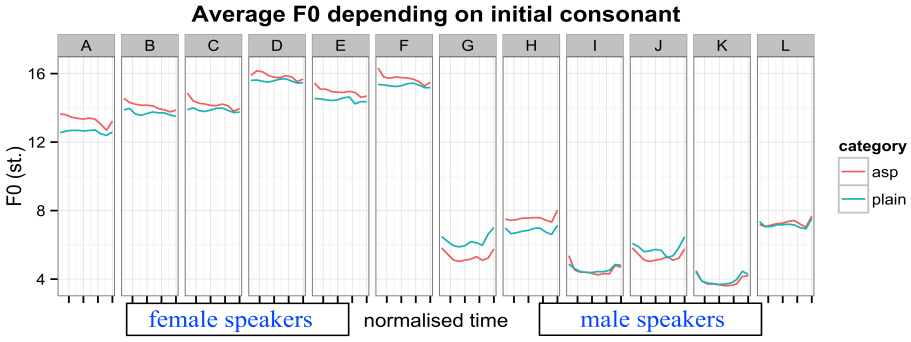


Figure 1: Time-normalised F0 contour of high level Fuzhou tone (6 females, 6 males)

Results suggest that Fuzhou Min is similar to Taiwanese Min in having higher average initial F0 following aspirated obstruents. This is true for both high level and high falling tones. Moreover, effects in the same direction, though on a smaller scale, are observed in these bilinguals' production of Mandarin words. This is in contrast to Beijing Mandarin speakers (Xu & Xu, 2003), showing that aspiration effect on F0 is language-specific and vary as a function of linguistic experience. Finally, female speakers consistently produce higher F0 following aspirated obstruents, while male speakers exhibit between-speaker variation, with speaker "G" and "J" showing the reverse pattern (See **Figure 1**).

The observed language- and gender-specificity in aspiration effect on F0 indicates that the phenomenon does not fall out from universal articulatory constraint alone; rather, it seems to vary with language and individual-specific laryngeal timing and configuration. While transglottal pressure difference has been offered as the driving force behind aspiration effects, this has been cited to explain two sets of diametrically opposed empirical findings in Xu & Xu (2003) and Lai et al. (2009). The status of the glottis may also play an important role: The outliers in the present study, speaker "G" and "J", seem to produce breathier vowels following aspirated obstruents, suggesting that wider glottal aperture following aspiration could cause pitch lowering, contrary to a previous claim in Xu & Xu (2003: 167). Diachronically, both aspiration-conditioned pitch lowering (Shuijingping Mang, Mortensen, 2013) and raising (Qingtian Wu, Steed, 2011) can be phonologised, showing that "aspiration", though a convenient cover term, may conceal detailed differences in laryngeal and supra-laryngeal coordination, which may variably affect F0 perturbation and provide diverse routes of phonologisation. Data collection on other varieties, including Beijing Mandarin, is ongoing.

The role of duration in the perception of prominence in typologically different and related languages

Yang Li (University of Cambridge), Ricky Chan (University of Cambridge), Adrian Leemann (University of Cambridge), Gerry Kwek (University of Cambridge), Anna Jespersen, (University of Cambridge), and Marie-José Kolly (University of Zurich)

Listeners of different languages have been reported to vary significantly in the perception of prominence patterns. Early examples of research in this area have shown, for instance, that Estonian listeners were more responsive to duration cues than English listeners [1]. These differences may be attributable to the greater role duration plays in the Estonian quantity system. Such findings suggest that the prosodic structure of listeners' native languages significantly influences their perception of prominence. To this day, there has been little systematic study, using unified procedure and stimuli, of the cues that are used in prominence perception for speakers of typologically different and related languages. In this contribution, we present a proof-of-concept study that attempts to fill this gap by examining the role of duration in prominence perception.

We used stimuli from [2] who studied the perception of prominence patterns in Standard German. [2] created the disyllabic logatome <baba>, which in its baseline form features identical durations in both syllables (stop duration was 63 ms, vowel duration (incl. burst) 189 ms). Based on this stimulus, [2] systematically modified the durations of the first and second syllable to create six stimuli: three stimuli with a stepwise increase in duration on the first syllable and likewise three manipulations on the second syllable. We used these stimuli in a binary forced choice task, where 80 listeners (8 varieties*10 listeners/variety) indicated which of the two syllables they perceived as stronger. Listeners were native-speakers of varieties that have been reported to differ in prosodic structure: Standard Southern British (SSBE) and Singapore (SgE) English, Valais (VS) and Schaffhausen (SH) Swiss German (SwG), Standard Danish, French, Hong Kong Cantonese, and Fuzhou.

Using baseline stimuli only (both syllables had identical durations), we found a significant effect of *variety* when comparing the relative proportions of responses (GLM, ($X^2(7)$=-14, p=.048*), see Figure 1 [3]:

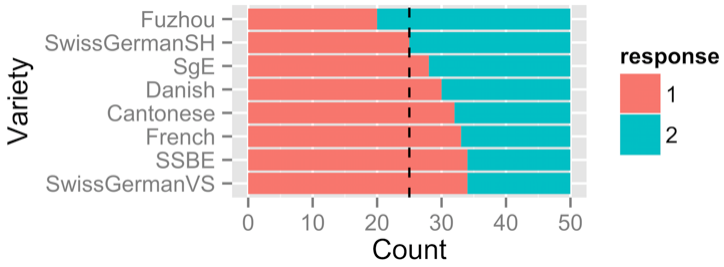


Figure 1: Responses by variety; perceived prominence on the first (red) or second (turquoise) syllable.

Figure 1 shows the relative proportions of responses to the first syllable (red) and second syllable (turquoise) for each variety (dashed line=50%). The most striking observation was that differences between varieties of the same language (e.g. SgE vs. SSBE or SwG SH vs. SwG VS) can be similar in magnitude to those found for typologically unrelated languages (e.g. Fuzhou vs. SwG VS). Potential explanations for these patterns may be found in the languages' diverse prosodic systems. The strong first syllable bias for SwG VS, for example, may have to do with its partly predictable first syllable stress, a relic from Old German stress rules. The difference between SgE and SSBE

may be explained with lower durational variability in vocalic intervals in SgE, which contributes to its syllable-timed rhythm [4]. This lack in durational variability may – amongst other factors – account for the listeners' lack of a bias towards perceiving prominence in the baseline stimulus. We also present data on the duration-manipulated stimuli. We found high sensitivity to duration cues in languages where durational cues play a primary role – or are important secondary cues – in marking contrast (e.g. SSBE [5]) and low sensitivity in languages that have been reported to lack such contrasts (e.g. French [6]). In future studies, we will further examine the contribution of pitch and intensity cues in the perception of prominence. It is conceivable that the relative contribution of these cues in prominence perception is a salient characteristic for between-language differences [7].

References

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

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

[3]A. Leemann, M.-J. Kolly, Y. Li, R. Chan, G. Kwek, A. Jespersen, "Towards a typology of prominence perception," *Proceedings of Speech Prosody 2016*, in review.

[4]E. L. Low, E. Grabe, and F. Nolan, "Quantitative characterisations of speech rhythm: syllable-timing in Singapore English," *Language and Speech*, vol. 43, pp. 377–401, 2000.

[5]P. Roach, "British English: Received Pronunciation," *Journal of the International Phonetic Association*, vol. 34, no. 2, pp. 239–245, 2004.

[6]S. Peperkamp, I. Vendelin, and E. Dupoux, "Perception of predictable stress: A cross-linguistic investigation," *Journal of Phonetics*, vol. 38, pp. 422–430, 2010.

[7]J. Vaissière, "Language-Independent Prosodic Features." in A. Cutler and D. R. Ladd (Eds.), *Prosody: Models and Measurement*. New York: Springer, pp. 53–66, 1983.

An articulatory investigation of Anglo-English prevocalic /r/

Natasha Lindley and Eleanor Lawson
Queen Margaret University, Edinburgh

To date, little work has been carried out regarding articulatory variation in Anglo-English prevocalic /r/ (but see (Delattre, Pierre & Freeman, Donald C. 1968). Anglo-English prevocalic /r/ is often transcribed as [ɹ], which is said to have labialization as a secondary articulation (see Hewlett and Beck 2006: 97, and West 1999). [ɹ̞] and other labial, variants, e.g. [ɹ̟̞], are noted as increasingly common in younger speech and in urban varieties in England (Foulkes and Docherty 2000; Docherty and Foulkes 2001). The phonetic relationship between [ɹ̈] with secondary labialization and primarily labial variants is unclear. It is also implied that labial variants have no specified tongue gesture; however, there is no study to date that investigates this idea. We report on an acoustic and articulatory study of prevocalic /r/ in seven varieties of Anglo English: Newcastle, S. Yorkshire, Manchester, Sheffield, London, southern Kent and Southampton.

Eight speakers aged 20 - 25 were selected from the *Dynamic Dialects* audio-ultrasound corpus (Lawson et al. 2015), recorded between 2011 and 2014. Thirteen words containing prevocalic /r/ were extracted from the corpus for each speaker. The words contained a word-initial singleton /r/, or /r/ in an initial cluster, followed by a range of vowels: /i/, /ɪ/, /ɛ/, /a/, /ʌ or ⋃/, /u/ and /aɪ/.

Tokens of /r/ were analysed both acoustically and articulatorily. Articulatory analysis involved extraction of midsagittal tongue surface contours at the point of maximum constriction for /r/, in order to identify characteristic tongue shapes involved in /r/ production. Lip protrusion was quantified using video from a headset-mounted profile lip camera. Maximum lip protrusion during the /r/ segment was measured in relation to neutral lip position before speech commenced. Acoustic analysis was carried out for F1 – F5, which were measured at the same temporal point as the articulatory point of maximum constriction.

Results show a continuum of tongue body shapes, from a tip up/retroflex tongue shape to a more vocalic tongue shape with a raised tongue dorsum and lowered tongue tip. One participant appeared to produce canonical labiodental /r/ with no specified tongue body position. One participant, who habitually produced a tip-up /r/, had secondary labiodentalisation on two tokens of /r/. Overall, an interesting correlation was observed between tongue body shape and lip protrusion; those speakers who used tip-up /r/ had the smallest degree of lip protrusion, while those with more vocalic tongue body shapes had the greatest degree of lip protrusion. Acoustically, tip up variants were associated with lower F3 minus F2 values (mainly due to lowering of F3) and higher F5 minus F4 values. There were no clear geographical patterns regarding either tongue shape variants or degree of lip protrusion.

These findings suggest that [ɹ] and [ɹ̞]/[ɹ̟̞] might not be as phonetically dissimilar as they first appear and that the relationship between lingual and labial [r] variants could be a phonetically gradient one, with intermediate articulatory variants containing specified, vocalic tongue-body shapes and stronger lip protrusion.

REFERENCES

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

Docherty, G.J. and Foulkes, P. 2001. Variation in (r) production: instrumental perspectives. *R-atics. Sociolinguistic, Phonetic and Phonological Characteristics of /r/. Etudes et Travaux,* 4 pp.173-184.

Foulkes, P. and Docherty, G.J. 2000. Another chapter in the story of /r/: 'Labiodental' variants in British English. *Journal of Sociolinguistics,* 4 (1) pp.30-59.

Hewlett, N. and Beck, J. 2006. *An Introduction to the Science of Phonetics.* London: Psychology Press.

Lawson, E., Stuart-Smith, J., Scobbie, J.M., Nakai, S., Beavan, D., Edmonds, F., Edmonds, I., Turk, A., Timmins, C., Beck, J., Esling, J., Leplatre, G., Cowen, S., Barras, W. and Durham, M. 2015. *Dynamic Dialects: an articulatory web resource for the study of accents.* Glasgow: University of Glasgow. Available at: http://dynamicdialects.ac.uk.

West, Paula ed. 1999. *The extent of coarticulation of English liquids: an acoustic and articulatory study. Proceedings of the International Congress of Phonetic Sciences.* , 1 - 7 August 1999,San Francisco.

The phonetics of questions asked in overlap

John Local (University of York) and Marianna Kaimaki (University of Cambridge)

There has been a considerable amount of research on the phonetic and interactional detail of overlapping talk. Researchers have documented the phonetic parameters people employ in designing competitive or non-competitive talk in overlap and have explored the phonetic design of talk which is recycled out of overlap. These studies have advanced significantly our understanding of how people use phonetic resources and phonetic variability to manage everyday conversation to resolve issues of speakership, establish themselves as the current speaker or make sure that information has reached their co-participant.

However, one particular interactional practice involving overlapping talk has not to date received systematic phonetic attention. There are sequences in which there is evidence to indicate that even when participants are engaged in talking themselves, they are also attending to and understanding another participant's co-occurring talk. These are sequences in which one speaker overlaps another and asks a question and, on exit from overlap, the question-recipient participant responds to and addresses the question that was co-produced with their own talk during the period of overlap.

In this paper, we examine all instances of questions asked in overlap in 18 hours of telephone calls which constitutes one-third of the CallHome American corpus (n=94). The dataset comprises 39 different phone calls involving 77 different speakers. The data reveal that in the majority of these instances overlapped speakers can and do respond unproblematically to the questions produced. The data also involve some instances where overlapped speakers do not respond to the questions in overlap but instead treat these questions as problematic and initiate interactional repair on them.

Through a detailed consideration of a range of pitch, loudness and duration measures and statistical modeling — Random Forests and normalized information gain Decision Trees — we seek to determine whether there are phonetic resources people use to make their talk as `hearable' (and unproblematic) or not by their co-participant. We show that, somewhat unexpectedly, questions in overlap which are treated as ‘heard' do not differ significantly in their phonetic design from those which are treated as problematic. Rather, the determining factor is the interactional sequence in which they are produced. We discuss the roles of speakers and hearers and explore some implications for existing models for speech production and perception.

An acoustic and articulatory study of syllable-medial glide insertion across China

Sarah Mahmood *University of Oxford* sarah.mahmood@ling-phil.ox.ac.uk

I present a study into a little known phenomenon: syllable-medial glide insertion in Mandarin Chinese. The motivation for this came from my own, and other people's perceptual reports. Syllable environments C_VN were studied, where V is a high vowel, either [i] or [u], and _ is the possible insertion of either [j] or [w]. The phonetic quality of the glides, as well as their distribution, both phonologically and geographically, were of interest.

Experiment I. Speech recordings from 33 participants (native speakers of Mandarin) were analysed with PRAAT 5.4.04 (Boersma & Weenink 2011). Randomized word-lists consisting of 150 tokens were used. The formants and intensity were analysed from the List A (44) tokens.

Known characteristics of glides include devoicing following voiceless stops and fricatives (Ohde & Scharf 1992: 83, 246), and frication. Both glides exhibit a narrow vocal tract constriction which lowers their frequency, though both also have more specific differences in terms of their formants, for example [w] should have a low F2 which falls below 1kHz, whereas [j] should have a high F2 that is closer to F3. Some glides have a steady-state period, in this case [w]. Thus, average formant measures of the glides and vowels were compared, with a further mid-point analysis conducted for [w] and [u] confirming the results of the average formant measures. Whether the glide is to be seen as part of the onset or the nucleus, may be attributable to differences in formant trajectories. Intensity measures revealed a surprising correlation between the intensity of the segments and the tone employed. It seems that for some speakers, whether the [j] or [i] is most intense depends on the tone, with T1 having a higher intensity vowel and T2 and T3 having a higher intensity glide. However, it should be noted that this pattern is not common across all speakers. Glide/vowel interaction is more complex than this, for example, in Korean Chinese; glides have been shown to have more energy than the vowel, therefore being a stronger contender for forming part of the nucleus of the syllable (Jin 2012: 96).

Experiment II. Articulatory data from 10 participants taken using the Ultrasonic Diagnostic Imaging System: Mindray DP-2200 Ultrasound machine (frame rate 60fps deinterlaced). Analysis was conducted with AAA software (Articulate Instruments 2010). Randomized minimal pairs word-lists were used. Specifically for glides, the tongue must be [+ATR] increasing volume in the pharyneal region, and forming a constriction near the soft or hard palate (Stevens 1998: 516).

Analysis & Discussion. Logistic mixed-effects modelling was conducted to test the effect of geographic location on glide insertion. [j]-insertion showed an almost categorical spread between north and south, and gradience in the central China group. [w]-insertion occurs across the three regions of China tested, not showing any significant influence of geographic location, though nevertheless variable in terms of how common it is for different speakers and regions. My analysis of the glide distribution patterns is based on two criteria: aspiration and place of articulation. The first criterion is adequate to explain the environment in which the northern speakers insert [j]: **[j] inserted before [i] iff Ons [+spread glottis]**. For the central group the environment is slightly more complex: **[j] inserted before [i] if Ons [+spread glottis] and/or Ons [+alveolar-palatal]**. The [w] distribution remains constant across the groups: **[w] inserted before [u] if Ons [+spread glottis]**. Hence, the phonological environments of [j]-insertion are: voiceless onsets [p^h, t^h, $tɕ^h$] in the north, [p^h, t^h, $tɕ^h$, $tɕ$, $ɕ$] in the central group, and again, only voiceless onsets for [w]-insertion.

I suggest that much of the phenomena detailed in these experiments stems from contrast maintenance. For example, when lacking a [w] glide, the need for contrast maintenance is gone, hence why we see F2 values return to the region of F2, for high vowels; normalizing within speaker. Contact bursts in the [j] spectrograms suggest a more consonantal articulation. Stevens (1998: 518) suggests that when [j] precedes high vowels, it is more constricted. Essentially, the constriction becomes narrower, and this could enhance the contrast between two relatively similar sounds. Furthermore, contrasting aspiration with frication as maximally distinct phenomena (Fant

1960: 169) in the case of [j] after voiceless sounds, and assuming that only segments which have a noticeable contrast appear beside one another, may be another method of achieving discrimination.

Authors/ Affiliation: Robert Mayr & Aysha Siddika (Cardiff Metropolitan University)

Title:

Inter-generational transmission in a minority language setting: Stop consonant production by Bangladeshi heritage children and adults.

Abstract:

A growing body of socio-phonetic work has examined the English accents of heritage language users, in particular in Asian minority language settings (e.g., Alam & Stuart-Smith, 2011; Evans et al. 2007; Heselwood & McChrystal, 2000; Kirkham, 2011; Kirkham & Wormald, 2015; Sharma & Sankaran, 2011). This work has shown substantial accentual differences between first-generation immigrants and their UK-born children with the latter being more affected by their peers' speech than their parents'. Few phonetic studies, have, however, examined changes to the *heritage language* in successive generations of speakers (but see McCarthy et al., 2013).

The present study aims to contribute to this latter line of inquiry by investigating the production of Sylheti and English plosives by two sets of Bangladeshi-heritage families from Cardiff (N=31). The first consists of first-generation female immigrants from the Sylhet area of Bangladesh and their British-born children (mean age: 4;5, range: 3;7-5;0), the second of second-generation British-born female heritage speakers of Sylheti and their (3^{rd} generation) children (mean age 4;3, range: 3;8-4;11).

All participants were recorded producing monosyllabic and bisyllabic words with word-initial plosives in Sylheti and English, using picture prompts. The Sylheti dataset included the categories, i.e. /p p^h b $b^ɦ$ t̪ ʈ ɖ̪ ɖ k k^h g $g^ɦ$/, the English one the categories /p b t d k g/. The data were analysed auditorily as well as acoustically.

The results revealed non-native productions of English plosives by the first-generation migrants but largely target-like patterns by the remaining sets of participants. The Sylheti stops exhibited incremental changes across successive generations of speakers, with the third-generation children's productions showing the greatest influence from English. While some of the patterns may be developmental in nature, the two sets of age-matched children showed systematic differences, indicating that input-related factors are at least equally important.

These findings have important implications for the maintenance, transmission and long-term survival of heritage languages, and show that investigations need to go beyond second-generation speakers, in particular in communities that do not see a steady influx of new migrants.

The Role of Formant Transitions in Identifying Plosives' Place of Articulation

Daniel McCarthy (Newcastle University)

How best to acoustically identify the place of articulation of plosives has attracted much attention and yet their identification remains far from fool-proof (e.g. Suchato's (2004) methods falsely classified the place of English plosives at a rate of 8%). Although there are two potential cues for identifying a plosive's place – its release burst and its formant transitions - the formant transitions have tended to be neglected by automatic speech recognition (see e.g. Ali et al., 2001: 838). Linked to this, ASR has normally found the burst to be a better cue than the transitions. But as one of these researchers concedes, '[…] it is possible that the inferiority of the formant-related attributes might be the result of our choice of acoustic measurements used to capture the formant information' (Suchato 2004: 161). In particular, correctly identifying alveolar stops from their formant transitions has been troublesome (e.g. in transition-only stimuli Suchato (2004) achieved 0% error in identifying labials, 5.9% error in velars, but a whopping 29.6% error in alveolars).

To address this problem, the present study compares the performance of some of the best-known methods for identifying place of articulation (Stevens et al., 1999; Suchato, 2004) with new acoustic formulae devised by the present author. The formulae compare the formant frequency of the vowel onsets and offsets with the frequencies of the vowels' midpoints and/or the peak frequency of the burst. The data consist of voiced stops in VCV nonce words uttered by two speakers of Received Pronunciation (one male, one female). The major challenge with using formant transitions as a cue is that they are a product both of the consonant *and* the surrounding vowels. Therefore the nonce words were designed to elicit as much variation in F2 as possible, i.e. V1, V2 = /iː ɛː ɜː ɑː ɔː/, C = /b d ɡ/, yielding 75 ˈVˈCV sequences (a format reminiscent of Öhman (1966)).

Preliminary results reveal that most of the formant-based formulae cannot correctly identify place of articulation to the same degree as existing methods. The relevance of these findings for deciding what identification algorithms to utilize in ASR is discussed. The results suggest that future research on how to identify plosives' place should focus on the burst as the primary cue.

References

Ali, A.M.A., Spiegel, J.V., and Mueller, P. (2001) 'Robust Classification of Consonants using auditory-based speech processing'. *International Conference on Acoustic, Speech, and Signal Processing*, Salt Lake City.

Öhman, S. (1966) 'Coarticulation in VCV utterances: Spectrographic Measurements'. *Journal of the Acoustical Society of America* 39(1), pp. 151-168.

Stevens, K.N., Manuel, S.Y., and Matthies, M.L. (1999) 'Revisiting Place of Articulation Measures for Stop Consonants: Implications for Models of Consonant Production'. *Proceedings of the 14th International Congress of Phonetic Sciences*, Volume 2, pp. 1117-1120.

Suchato, A. (2004) '*Classification of Stop Consonant Place of Articulation*'. PhD Thesis, Massachusetts Institute of Technology.

Stress realisation in Sylheti-English bilinguals from the London Bengali community

Kathleen M. McCarthy[1] and Esther de Leeuw[2]
[1]University College London; [2]Queen Mary University of London

The aim of the current study was to examine the influence of language background and language use patterns on first (L1) and second language (L2) lexical stress realisation in Sylheti-English bilinguals from the London Bengali community. In English, a primary acoustic cue of lexical stress realisation is a higher pitch in the stressed syllable in relation to the neighbouring unstressed syllables (Reetz & Jongman, 2009). Although very little is known about Sylheti, it has been suggested that the opposite holds true (see Reetz & Jongman, 2009; Hayes & Lahiri, 1991).

Twenty-seven Sylheti-English bilinguals from the London Bengali community took part in the study. Speakers were divided into three groups: 1) Late bilinguals (*n*=11, mean AoA=21 yrs); 2) Early bilinguals (*n*=9, mean AoA=6 yrs); 3) Second-generation bilinguals (*n*=6), born in London. Twelve control speakers were recorded in London: Standard Southern British English (SSBE) monolingual speakers (*n*=6), and Sylheti native speakers (*n*=6), the latter of which had arrived in the UK a maximum of three months before the recording and had received no formal English training.

Using a picture naming task, speakers were recorded producing disyllabic Sylheti and English words with penultimate syllable stress in the carrier sentences "abar ____ kho" and "Say ____ again". Maximum F0 measurements were extracted for the penultimate and ultimate vowel in the target words (Mennen, Mayr and Morris, 2015). The difference between maximum F0 of the penultimate and ultimate vowel was expressed in semitones, such that a positive value indicated a rise from the penultimate syllable to the ultimate syllable (i.e. predicted for Sylheti monolingual speech), and a negative value indicated a fall from the penultimate syllable to the ultimate syllable (i.e. predicted for English monolingual speech). We predicted that the bilinguals would vary in their production pattern dependent on their use of Sylheti and SSBE.

Our preliminary results (see table 1) indicated that the Sylheti monolinguals realised lexical stress in the penultimate syllable of the target word by means of a lower F0 than in the ultimate syllable, whilst in the English monolinguals the opposite was the case. The bilinguals' Sylheti and English stress pattern varied dependent on their language background. Specifically, the late bilinguals realised Sylheti and English stress by means of a slight rise from the penultimate syllable, approximating the Sylheti-like stress pattern in both English and Sylheti. In contrast, the second-generation bilinguals realised Sylheti and English stress by a fall in F0, approximating the English stress pattern. Interestingly, the early bilinguals displayed a distinct pattern for English and Sylheti, in the direction of the monolingual patterns. A further durational analysis is planned to determine how different acoustic cues affect lexical stress realisation in the bilinguals. These findings have implications for our understanding of how language background and environment influences L1 and L2 language acquisition.

Table 1. Speakers' pitch realisation in semitones

	Sylheti		English	
	Mean	SD	Mean	SD
Sylheti control	1.60	2.36	n/a	n/a
Late bilinguals	0.61	1.61	0.79	1.36
Early bilinguals	0.21	1.55	-0.58	1.58
2nd gen bilinguals	-0.17	1.07	-0.99	1.92
English control	n/a	n/a	-1.30	1.72

The influence of language exposure and social networks on bilingual phonemic development

Kathleen M. McCarthy and Bronwen G. Evans
University College London

Previous research on bilingual language development has shown that the amount of language exposure influences children's speech and language outcomes in each of the languages being learned. In general, findings indicate that children display better outcomes for vocabulary, grammar, and more recently for speech perception in their dominant language (e.g., De Houwer, 2009; Garcia-Sierra et al., 2011; Hoff 2012; Pearson et al, 1997; Place & Hoff, 2011). To date, this research has focused on simultaneous bilinguals who are exposed to both of their languages from birth. Yet, bilingual children who grow up in immigrant communities have a more complex language environment. Typically, these children are exposed to their home language until they enter full time education, when they are immersed in English. These children's exposure patterns are greatly determined by the language background of the children's social network ties. For example, the input a child receives from first generation grandparent is likely to differ from their second-generation older sibling (see e.g., McCarthy et al., 2013). The aim of the current study was to investigate the influence of language exposure and social networks on the English speech perception and production skills of Sylheti-English sequential bilingual children from the London Bengali community.

To investigate the influence of these factors a social network approach was adopted (see Wei, 1992; Sharma, 2011). Thirty-eight Sylheti-English children (4-5 years old) and their main caregiver took part in the study. Data regarding children's regular interlocutors (network ties) was collected during caregiver interviews. For each network tie the following information was collected: the language/s used, the amount of time spent with the child in hours, the context of the interaction and the speaker's background e.g., age of arrival in the UK. Two analyses were conducted. For language exposure, children were grouped according to the amount of English exposure they received before and after entering full-time education. These groups were subsequently compared on their English speech perception (VOT categorization, vowel identification) and production measures (acoustic analysis of stop and vowel production). Correlational analyses were conducted to investigate the relationship between the children's network characteristics (i.e., background of network members) and their speech perception and production outcomes.

In line with previous bilingual research our results showed that language exposure was related to children's speech perception and production development. The children who had less English exposure had significantly lower vowel ID scores and shallower voicing categorization slopes than those with high English exposure. For production, children with less English exposure used a shorter VOT than did those with high English exposure. No difference between the exposure groups was found for vowel production. For social networks, the correlation analyses showed that some but not all factors were related to the children's outcome measures. In particular, sibling language use was significantly correlated with the children's perception and production. Furthermore, the proportion of first generation Bangladeshi network ties was related to the production of English back vowels i.e., FOOT and GOOSE fronting. These findings have implications for our understanding of the relationship between linguistic input and the phonemic development of bilinguals in complex linguistic environments.

Listeners' judgements of the similarity of voices with different accents

Kirsty McDougall, University of Hertfordshire and University of Cambridge
Toby Hudson, University of Cambridge
Nathan Atkinson, University of York
kem37@cam.ac.uk, toh22@cam.ac.uk, nathan.atkinson@york.ac.uk

'Perceived voice similarity' (PVS) is the notion that within a group of speakers of the same sex, age and accent background, listeners will perceive certain speakers as sounding more similar to each other than others. This is due to the individual differences in vocal tract anatomy and physiology present among the group of speakers combined with the individual choices (whether conscious or subconscious) they make in implementing their linguistic systems. This paper reports part of an ongoing research programme which explores the perceptual workings of this individual variation. The present paper investigates the role played by accent in judgements of PVS, in particular how similar listeners judge speakers of different accents and of the same accent to sound.

The paper builds on an earlier study in which separate experiments were conducted for two accents, Standard Southern British English (SSBE) and York English (YE). Stimuli were constructed using recordings of male speakers aged 18-25 years: 15 SSBE speakers from the *DyViS* database (Nolan *et al.* 2009) and 15 YE speakers from the *YorViS* database (McDougall *et al.* 2015). Two utterances of spontaneous speech per speaker, each approximately 3 seconds in duration, were excised from a telephone call task. In each experiment, two listener groups (SSBE-speaking and YE-speaking) rated the (dis)similarity of a set of pairings of speakers for one of the accents on a 9-point Likert scale. Listener judgements were subjected to Multidimensional Scaling (MDS), a data reduction technique which enables a group of speakers to be characterised in a perceptual space whose dimensions can be further interpreted to better understand the relationships of PVS among each group of speakers. For each accent, the perceptual dimensions resulting from the MDS were correlated against measurements of phonetic variables (long-term fundamental frequency, articulation rate and long-term formant analysis of F1, F2, F3), and results for the two listener groups compared. Correlation results showed different phonetic features contributing to PVS judgements for the two accents. Long-term fundamental frequency is of key importance for PVS in SSBE (for both SSBE and YE listeners), but does not show significant correlation with any of the perceptual dimensions generated by the YE stimuli. Articulation rate shows some importance for PVS in YE for SSBE-speaking listeners only, but is not significant in SSBE. Long-term formants are significant in both accents: F1 and F4 for SSBE and F2 and F3 for YE, again with differences between the two listener groups.

Results will be presented for a further experiment in which listeners judged the (dis)similarity of a group containing half SSBE speakers and half YE speakers. The stimuli from the previous experiment were used in an experiment of the same format, for which seven speakers were chosen at random for each of the two accents. All possible speaker-pairings were included, such that the listeners judged the similarity of pairs of SSBE speakers, pairs of YE speakers and pairings of an SSBE and a YE speaker within the same task, presented in a random order. 40 listeners took part: 20 speakers of SSBE and 20 speakers of YE. MDS and correlation analysis will be used to evaluate the relative roles of acoustic-phonetic features such as fundamental frequency, articulation rate and long-term formant frequencies, and of the speakers' accents. Implications of the findings for a phonetic theory of perceived voice similarity and for foil selection in the practical application of voice parade construction will be discussed.

References

McDougall, K., M. Duckworth and T. Hudson (2015) 'Individual and group variation in disfluency features: a cross-accent investigation.' In: The Scottish Consortium for ICPhS 2015 (ed.)

Proceedings of the 18th International Congress of Phonetic Sciences, 10-14 August 2014, Glasgow. Paper number 0308.1-4. <http://www.icphs.info/pdfs/Papers/ICPHS0308.pdf>

Nolan, F., K. McDougall, G. de Jong and T. Hudson (2009) 'The *DyViS* database: style-controlled recordings of 100 homogeneous speakers for forensic phonetic research', *International Journal of Speech, Language & the Law* 16(1): 31-57.

Transfer and changing linguistic norms: the rhotic consonants of Bearnese Gascon

Damien Mooney (University of Bristol)

The southern Gallo-Romance (or *langue d'oc*) variety historically spoken in the region of Béarn in southwestern France is commonly referred to as 'Béarnais' (or 'Bearnese'), a sub-dialect of the Gascon language. From the late 19th century onwards, and indeed over the course of the 20th century, Bearnese has found itself in an increasing state of language obsolescence, contracting under pressure from the dominant French language and exhibiting contact-induced linguistic changes at all levels of linguistic structure.

The phonological inventory of Bearnese contains two rhotic consonants, the voiced apical tap /ɾ/ and the voiced apical trill /r/ (Grosclaude, 1986 : 4). Traditonally, these phonemes are only found to be in constrastive distribution in intervocalic position, e.g. *poret* /puˈɾet/ ('chicken') ~ *porret* /puˈret/ ('leek'); in some Gascon varieties, it has been suggested that tapped and trilled rhotics are in complementary distribution in non-intervocalic contexts, such that 'an archiphoneme could be set up for all other positions' (Cardaillac Kelly, 1973: 32). The distribution of [r] and [ɾ] is somewhat constrained by their position within the syllable and with respect to word boundaries with a tendency for [r] to occur word-initially and as an onset after [n], and for [ɾ] to occur in tautosyllabic onset clusters, as an onset after consonants other than [n], and in the syllable coda (Cardaillac 1973: 32), but this distribution is by no means categorical (Mooney, 2014: 345).

Over the course of the twentieth century, Bearnese has adopted some dorsal rhotic consonants as a result of prolonged language contact with French. The phonological inventory of modern standard French contains one rhotic consonant phoneme, the voiced uvular fricative /ʁ/, which is often realised as a voiced uvular trill [ʀ] by older rural speakers. Additionally, there is evidence to suggest that /ʁ/ is being realised increasing as a voiced uvular approximant, particularly in final position (Fougeron and Smith, 1999: 80). It is worth noting that both [r] and [ɾ] occurred in French until the eighteenth century but by the time of the Revolution (1789), in Paris at least, they had largely been replaced by the uvular fricative [ʁ] (Rickard, 1974: 108).

In the Gascon of Donzac, north of Béarn, Cardaillac Kelly found that the historically appropriate phonemes, /ɾ/ and /r/, were realised variably as dorsal rhotics, and in particular as [ʀ], 'in *all* positions as a consequence of bilingualism' (1973: 32), noting that the adoption of French [ʀ] for both /ɾ/ and /r/ in intervocalic position may result in the neutralisation of phonemic contrast in Gascon. These findings suggest that contact-induced dorsal variants find themselves in free variation with both apical variants and that there are no linguistic constraints on the transfer involved. Female speakers were shown, however, to be leading this contact-induced change, adopting substantially more dorsal variants into their variety of Gascon than male speakers (Cardaillac Kelly, 1973: 32).

This presentation will examine empirically the distribution of apical rhotics in Bearnese Gascon, as well as evidence for contact-induced linguistic change from French. Importantly, the linguistic and social constraints on change will be considered: evidence for language-internal constraints on the adoption of dorsal rhotics will be advanced; the hypothesis that the speech of female speakers is characterised by higher levels of transfer will be tested. Acoustic and auditory data from a wordlist translation task undertaken with 10 bilingual Bearnese-French speakers (5 male; 5 female) will be presented (400 tokens in total), focusing on the phonetic and phonological distribution of apical and dorsal rhotics in four phonological environments: simple onsets (σ_V), complex onsets (σC_V), simple codas (V_σ), complex codas (VC_σ). For each variant, the analysis will consider its place of articulation (apical; dorsal), manner of articulation (tap; trill; fricative), and whether the segment is fully voiced or devoiced. A qualitative analysis of the acoustic waveform and spectrogram will be used to classify each rhotic consonant, e.g. repeated cycles of low amplitude for trilled variants, aperiodic waveform and amplitude concentrated at high frequencies for fricatives, etc. A mixed-effects logistic regression analysis in Rbrul (Johnson, 2008), including 'speaker' and 'word' as random effects, will consider the effect of a number of fixed-effect predictors on rhotic voicing, place of articulation, and manner of articulation: syllable type; stress; preceding and following phonological environment; speaker sex. The linguistic and social mechanisms governing transfer in Bearnese Gascon will be discussed fully in order to account for the distribution of historically appropriate and contact-induced rhotic consonants in this highly obsolescent variety of southern Gallo-Romance.

References: CARDAILLAC, K. R. 1973. *A Descriptive Analysis of Gascon* (The Hague and Paris: Mouton)**;** FOUGERON, C., and C. L. SMITH. 1999. 'Illustrations of the IPA: French', *Handbook of the International Phonetic Association* (Cambridge: Cambridge University Press), pp. 78–81; GROSCLAUDE, M. 1986. *La langue béarnaise et son histoire: étude sur l'évolution de l'occitan du Béarn* (Orthez: Per Noste); JOHNSON, D. E. 2008. 'Rbrul' <http://www.ling.upenn.edu/~johnson4/Rbrul.R>; MOONEY, D. 2014. 'Illustrations of the IPA: Béarnais (Gascon)', *Journal of the International Phonetic Association*, 44: 343–350**;** RICKARD, P. 1974. *A History of the French Language* (Oxford: Routledge).

How speech sounds could be intrinsically perceptuo-motor

Piers Messum
Pronunciation Science Ltd, London

There has been a long-standing debate within phonetics about the neurophysiological status of speech sounds, including the contention of Sapir (1921) that the auditory aspect of speech was primary and the contrary view of Stetson (1951) that speech is "gestures made audible." Most recent contributions characterise the cognitive nature of phonological units as 'perceptuo-motor' but are actually aligned with Sapir's view in that the motor component is associated with an auditory primitive rather than being of equal status with it.

Speech sounds are the outcome of two practical processes: a child learning to (1) comprehend and (2) pronounce a particular language. Basic assumptions about how these take place set the terms of the debate about the neurophysiology. In particular, in the learning of pronunciation it has always been assumed that the child solves the correspondence problem for speech sounds by imitation. That is, the child uses his own judgement of similarity between what he recovers from the speech input (an acoustic pattern in some theories or a pattern of gestures in others) and what he produces in return to match this. This judgement informs and improves his subsequent production of speech sounds in a 'matching to target' process. A judgement of similarity is only possible between images that are comparable, so the child is assumed to be operating either with auditory or motor primitives, and the favoured one of these is then considered to be the form for the underlying representation of speech in the human brain.

There is, however, no evidence that young children solve the correspondence problem by imitation (by self-supervised auditory matching). Correspondence problems can also be solved by the mirroring behaviour of social partners, who reflect information about the behaviour of the learner back to him. The well-documented, ubiquitous phenomenon of the reformulation of an infant's utterances into L1 speech sounds by his caregivers is an example of such behaviour. When this takes place in response to the output of a vocal motor scheme (a well-practised, stable motor routine), simple association is all that is then necessary for the correspondence problem to be solved by the infant, since there will be correlated excitation of his sensory and motor neurons (Cook et al. 2014).

Learnt in this way, speech sounds would not be fundamentally auditory or fundamentally motor, but would be intrinsically perceptuo-motor, matching the nature of speech itself: motor in production, auditory in perception. I will discuss the evidence for this account and the anomalies it resolves. Further detail can be found in Messum and Howard (2015).

Cook, R. et al. (2014). Mirror neurons: From origin to function. *BBS*, *37*, 177–192

Messum, P. R., & Howard, I. S. (2015). Creating the cognitive form of phonological units: The speech sound correspondence problem in infancy could be solved by mirrored vocal interactions rather than by imitation. *Journal of Phonetics*, 53, 125–140.

Sapir, E. (1921). *Language*. New York: Harcourt, Brace and World

Stetson, R. H. (1951). *Motor Phonetics*. Amsterdam: North Holland.

Linguistic and extra-linguistic constraints on preaspiration in Bethesda Welsh

Jonathan Morris (University of Cardiff) and Michaela Hejná (Newcastle University)

This paper examines three dimensions of preaspiration (frequency of occurrence, duration, and noisiness) and the language-internal and sociolinguistic factors which influence its production in the Welsh of Bethesda (North West Wales). Firstly, the paper contributes to our knowledge of preaspiration as a feature of Welsh. Secondly, it compares the role of linguistic factors with previous research in other languages and, in particular, evaluates the effects of speaker sex and stop type. Thirdly, it examines the effect of differing modes of acquisition of Welsh (parental transmission vs. immersion in the community and school) on speech production and variation in a bilingual community.

Ball (1984: 18) suggests that preaspiration (normally defined as aspiration preceding voiceless segments; Laver, 1994: 356) could be a feature of Welsh. Since then, preliminary studies have shown that it is variably present both in the English and Welsh of northern and central areas of Wales (Morris, 2010; Hejná, 2015) and in central and southern areas of the country (see Hejná, 2015 for English and Iosad, in press for Welsh). There have been no systematic investigations of linguistic and extra-linguistic influences on its production in Welsh and none which consider frequency of occurrence, duration, and noisiness.

The study is based on wordlist data from 16 Welsh-English bilinguals aged between 16 and 18, eight of whom use Welsh and eight of whom use English as their home language. Within each home-language group, there are four female and four male speakers. The wordlist contains word-medial and word-medial instances of fortis (n=698) and lenis (n=266) plosives, which are generally phonetically voiceless. The data were analysed acoustically using Praat (Boersma & Weenink, 2015). Preaspiration was coded as present if a period of glottal frication preceded the stop. Durational data (where applicable) were obtained using a script after manual segmentation and log-transformed. We used band-pass filtered zero-crossing rate (BP ZCR) data, obtained via script (Gordeeva & Scobbie, 2010), as a measure of noisiness. Data were analysed statistically using mixed-effects regression modeling in R (R Core Team, 2015) with word and speaker as random factors.

Preaspiration was found in 87% of the fortis stops and 48% of the lenis stops. Speaker sex and home language were not found to be significant in the fortis series, but females pre-aspirate more frequently in the lenis series. Speaker sex, however, was found to be a significant predictor of duration and noisiness with females producing longer and noisier preaspirated tokens both in the fortis and the lenis series. Home language was found to influence the duration of pre-aspiration in the fortis series, with speakers using Welsh at home having longer durations. Vowel height and place of articulation were also found to be significant predictors.

The results indicate that preaspiration is a variable feature in Bethesda Welsh in general, but that there are phonetic differences in its production between males and females. Interestingly, however, the linguistic factors (namely vowel height and place of articulation) exert a greater influence on the feature than speaker sex. This, we argue, suggests that universal differences between stop articulation are more important than possible physiological differences between males and females. The fact that home language was not significant for most aspects of preaspiration will be discussed with reference to recent studies (e.g. Morris, 2013; Mennen et al., 2015; Mayr et al., forthcoming) which have found no significant effect of mode of acquisition at the level of phonetics but socially-conditioned phonological and supra-segmental variation.

How do we distinguish between read and spontaneous speech?

Rosanna Morris-Haynes[1], Laurence White[1], Sven L. Mattys[2]

1School of Psychology, Plymouth University, 2Department of Psychology, University of York.
rosanna.morris-haynes@plymouth.ac.uk; laurence.white@plymouth.ac.uk;
sven.mattys@york.ac.uk

Speech rate, mean pitch, pitch range and boundary marking cues have been found to differ between read and spontaneous speech, and to contribute to listeners' discrimination of the two speech styles (e.g., Batliner, Kompe, Kiessling, Nöth & Niemann, 1995; Blauw, 1994). However, some studies have found poor speech style discrimination by listeners, despite the availability of reliable cues, including more frequent utterance-final pitch rises in spontaneous speech (e.g., Dellwo, Leemann & Kolly, 2015; Morris-Haynes, White & Mattys, 2015).

Using spontaneous map-task utterances and lexically-identical read counterparts extracted from an existing corpus, we first asked whether the opportunity to make a direct pairwise comparison between utterances in read/spontaneous pairs would facilitate recognition of the final pitch movement cue. Secondly, we tested whether prior participation in a map task – comparable to that used to elicit the spontaneous speech stimuli – would facilitate listeners' identification and interpretation of cues to speech style.

We found that, whilst direct comparison between paired utterances did not in itself boost listeners' performance, the priming provided by prior participation in a map task did facilitate style discrimination. Thus, even when prosodic cues to speech style are reliably available, listeners' interpretation of features such as rising final pitch varies according to contextual factors. Increased access to the production context of comparable speech may be beneficial for style discrimination in that it enables listeners to more accurately map prosodic form to function.

References

Batliner, A., Kompe, R., Kieβling, Nöth, E., & Niemann, H. (1995). Can you tell apart spontaneous and read speech if you just look at prosody? *Speech Recognition and Coding- New Adventures and Trends* (pp. 101-104).

Blauuw, E. (1994). The contribution of prosodic boundary markers to the perceptual difference between read and spontaneous speech. *Speech Communication*, *14*, 359-375

Dellwo, V., Leemann, A., & Kolly, M.-J. (2015). The recognition of read and spontaneous speech in local vernacular: The case of Zurich German. *Journal of Phonetics*, *48*, 13-28.

Morris-Haynes, R., White, L., & Mattys, S.L. (2015). What do we expect spontaneous speech to sound like? In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 16th International Congress of the Phonetic Sciences*. Paper 1011.

Trends and Future of Web-Based Articulatory Video Resources

Satsuki Nakai[1], David Beavan[2], Eleanor Lawson[1], Grégory Leplâtre[3], James M. Scobbie[1], and Jane Stuart-Smith[4]

[1] Queen Margaret University, [2] University College London, [3] Edinburgh Napier University, [4] University of Glasgow

Speech production is one of the most complex motor activities performed by humans, involving delicately coordinated, multiple articulators (e.g., vocal cords, lips, velum, tongue, jaw). Many of these articulators are difficult to see during speech production. Thanks to advances in technology, an increasing number of videos of hidden articulators have become available online in recent years, enabling anyone with Internet access to observe speech articulators in action for free. This paper will (1) introduce recent developments in imaging and web technologies and some web-based resources that showcase such developments; and (2) discuss the future of web-based articulatory video resources.

Techniques to image hidden speech articulators have been dramatically improved in the past decade. Notably, several research groups (e.g., Narayanan et al., 2014) have developed new data acquisition and image reconstruction techniques to address the issue of slow frame rate of magnetic resonance imaging (MRI). Such state-of-the-art MRI videos (at 83 fps) can be viewed on *rtMRI IPA Chart* hosted by the University of Southern California. Others (e.g., Lawson et al., 2015) have combined MRI data, which can show the entire vocal tract, with temporally-refined data of the tongue obtained using UTI (ultrasound tongue imaging). Animated heads created using such a technique can be found on *Seeing Speech* hosted by the University of Glasgow. Other recent developments in imaging technologies, which have implications for future resources, include dynamic 3D visualisation of the vocal tract using MRI (e.g., Zhu, et al., 2013) and a technique to image teeth in MRI (Idiyatullin et al., 2011).

Web technologies are also moving very fast (if not faster), driven, in part, by the rapid adoption of smartphones and tablets. For example, the video delivery method is moving from plugins such as QuickTime and Adobe Flash to HTML5, a new international web standard, which allows video playback on smartphones and tablets as well as computers. Both *rtMRI IPA Chart* and *Seeing Speech* are among the websites designed to cater for audience who wish to view articulatory videos without a plugin. Modern devices like smartphones and tablets are also bringing new immersive ways to interact with the videos, with touch-screen interfaces and sensors to track orientation. These abilities of the modern devices are expected to facilitate 3D representations of speech articulators.

References

Idiyatullin, D., Corum, C., Moeller, S., Prasad, H. S., Garwood, M., & Nixdorf, D. R. (2011). Dental magnetic resonance imaging: Making the invisible visible. *Journal of Endodontics, 37*(6), 745–752.

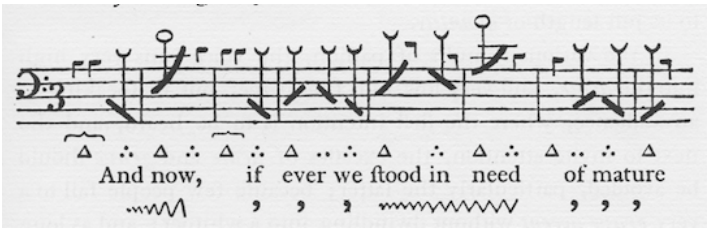
Narayanan, S., et al. (2014). Real-time magnetic resonance imaging and electromagnetic articulography database for speech production research (TC). *Journal of the Acoustical Society of America, 136*(3), 1307–1311.

Zhu, Y., Kim, Y.-C., Proctor, M. I., Narayanan, S. S., & Nayak, K. S. (2013). Dynamic 3-D visualization of vocal tract shaping during speech. *Medical Imaging, IEEE Transactions on*, *32*(5), 838–848.

The rise and fall of the British School of intonation

Francis Nolan
Phonetics Laboratory, DTAL, University of Cambridge

This paper will give a selective history of the evolution of the 'British' framework for intonation analysis from Steele (1775) [extract right] to its zenith in the mid twentieth century. The tour will call at landmarks such as Sweet (1877), Jones (1909), Palmer (1922), and Kingdon (1958). I will also assess its role and scientific value in comparison to the autosegmental-metrical (AM) framework dominant in intonation research today.



One of the themes will be the faltering but ultimately triumphant emergence of a phonological approach to intonation, one which looks for abstract contrastive units having both a variable relationship to functions and context-dependent realisations, orders them in paradigms with respect to slots in an internal structure accorded to the intonational phrase, and seeks co-occurrence restrictions between them. I will argue that at its best the British School, though short on quantitative instrumental data, showed phonological sophistication at least as great as evident in the AM enterprise. As for adjudication between analyses in the British framework, with its reliance on pitch movements, and AM with its H and L targets or turning points, I will question whether alignment data incontrovertibly favour a H/L target approach as has been claimed (e.g. Arvaniti *et al.* 1998).

Overall, despite the superficial dissimilarities between the frameworks there is more common ground than might at first be expected. The IViE model within the AM framework was an early attempt to exploit the commonalities and form a bridge. But, with an eye to the overwhelmingly pedagogical aims of most British analyses, I will suggest that the adoption of one concept from AM, the boundary tone, could increase the transparency and learnability of intonation systems represented in a broadly British framework using iconic symbols. Conversely, the case will put for the British School's treatment of syllables made prominent without pitch change. In conclusion, it will be suggested that the time is ripe for a re-evaluation of the British approach, backed up by quantitative empirical research.

Arvaniti, A., Ladd, D.R. and Mennen, I. (1998) Stability of tonal alignment: the case of Greek prenuclear accents. *Journal of Phonetics* 26, 3–25.

Jones, D. (1909) *Intonation curves.* Leipzig and Berlin: Teubner.

Kingdon, R. (1958). *The groundwork of English intonation.* London: Longman.

Palmer, H.E. (1922). *English intonation, with systematic exercises.* Cambridge: Heffer.

Steele, J. (1775) *An essay towards establishing the melody and measure of speech, to be expressed and perpetuated by certain symbols.* Facsimile edition (1969), Menston, UK: The Scolar Press Limited.

Sweet, H. (1877, 2013) *A handbook of phonetics : including a popular exposition of the principles of spelling reform.* Oxford: Clarendon.

Coding categories relevant for interaction

Richard Ogden (University of York)

This contribution looks at how speech shapes, and is shaped by, interaction. Starting with a brief overview of how interactionally relevant categories have already been coded, we look at the implementation and organisation of social actions through talk, focusing on the use of clicks in spoken English conversations. The aim of the paper is to highlight the relevance of interactional categories in sociophonetic analysis.

Clicks in English are have several distinct, partially conventionalised, forms in English: they can be single or multiple, central or lateral, and they can be nasalised. They occur in turn initial, medial or final position, and this positioning contributes to their interpretation. Clicks are a feature of interaction, and have no lexical function in English; so a coding scheme is required that considers their interactional functions alongside their phonetic forms and distributions in turns at talk. While there is rather little video data on clicks, the material that we have suggests that visual features also contribute significantly to their function: body posture, facial expression and gesture are features of their composition. It has been shown (Ogden 2012) that on the face of it there is a lot of individual variation in the usage of clicks: some speakers apparently produce more than others, and it looks like female speakers produce more clicks on average than male speakers. However tempting it might be to think of clicks as indexical, this finding takes no account of the functions of clicks. If clicks are used in the delivery of some social action, then their distribution is accounted for by that social action. This in turn begs the question of how social actions are implemented through talk, and any coding scheme needs to handle that.

This paper is based on a large collection of clicks in spoken British and American English audio data. One of the primary functions of clicks is to mark displays of affective stance. Looking closely at a case study, and comparing it to other data from the corpus, and to published data in affective displays in conversation, we will consider how displays of affective stance are made relevant and implemented in conversation, with a focus on phonetic design features of such turns. We will argue that a coding scheme for interaction needs to handle the dialogic nature of interaction, and that alongside information about speakers' identities, coding schemes also need to recognise the social actions conveyed through talk moment by moment, which must be based on a close analysis of the talk itself, and not on broad aspects of its genre.

Patterns of Articulatory Activation in Delayed Naming

Pertti Palo, Sonja Schaeffler and James M. Scobbie

Clinical Audiology, Speech and Language (CASL) Research Centre, Queen Margaret University

We study the effect that phonetic onset of an utterance has on acoustic and articulatory reaction times (RTs). An acoustic study by Rastle et al. (2005) shows that the place and manner of the first consonant or consonant cluster in a target affects acoustic RT. An articulatory study by Kawamoto et al. (2008) shows that the same effect is not present in articulatory reaction time of the lips.

In this study we answer the question: Can we predict the location of the first tongue movement with the phonological identity of the utterance onset? Knowing the phonological features of the initial consonantal segment one might think that it is easy to predict that certain structures are going to move first. However, from a coarticulatory point of view, we know that what happens first might actually be related to the production of the second, third or even fifth segment.

To answer our research question we analyse tongue ultrasound data from a delayed naming experiment using 58 lexical English words. The words all have a high lexical frequency and are of the types /VC/, /CVC/, /CCVC/, and /CCCVC/. The consonants and vowels are varied systematically in the corpus. We will present data from three Standard Scottish English spekears. The analysis technique is the subject of another submission to BAAP 2016 by the same authors.

References

Kawamoto, A. H., Liu, Q., Mura, K., and Sanchez, A. (2008). Articulatory preparation in the delayed naming task. *Journal of Memory and Language*, 58(2):347 – 365.

Rastle, K., Harrington, J. M., Croot, K. P., and Coltheart, M. (2005). Characterizing the motor execution stage of speech production: Consonantal effects on delayed naming latency and onset duration. *Journal of Experimental Psychology: Human Perception and Performance*, 31(5):1083 – 1095.

Identifying the First Structure to Move in Ultrasound Videos

Pertti Palo, Sonja Schaeffler and James M. Scobbie

Clinical Audiology, Speech and Language (CASL) Research Centre, Queen Margaret University

We study the effect that phonetic onset of an utterance has on acoustic and articulatory reaction times (RTs). An acoustic study by Rastle et al. (2005) shows that the place and manner of the first consonant or consonant cluster in a target affects acoustic RT. Kawamoto et al. (2008) have shown that the same effect is not present in articulatory RT of the lips.

In this study we seek to identify the articulatory structures that move first in each context. We do this by developing an automated analysis method for ultrasound tongue videos. Ultrasound data in its raw form consists of scanlines (Figure 1). The analysis is based on calculating a Euclidean distance measure for each scanline of the raw ultrasound data. Similar methods have been employed as holistic measures by McMillan and Corley (2010) and Raeesy et al. (2011).

The location of the first movement is identified by finding the scanline with maximal change at movement onset. Movement onset is identified initially by manual inspection of the ultrasound videos but validated by the scanline based measure by checking that the ultrasound sequence does not contain clear change maxima before the identified onset time. The results from analysing delayed naming data from a group of participants is the subject of another submission to BAAP 2016 by the same authors.

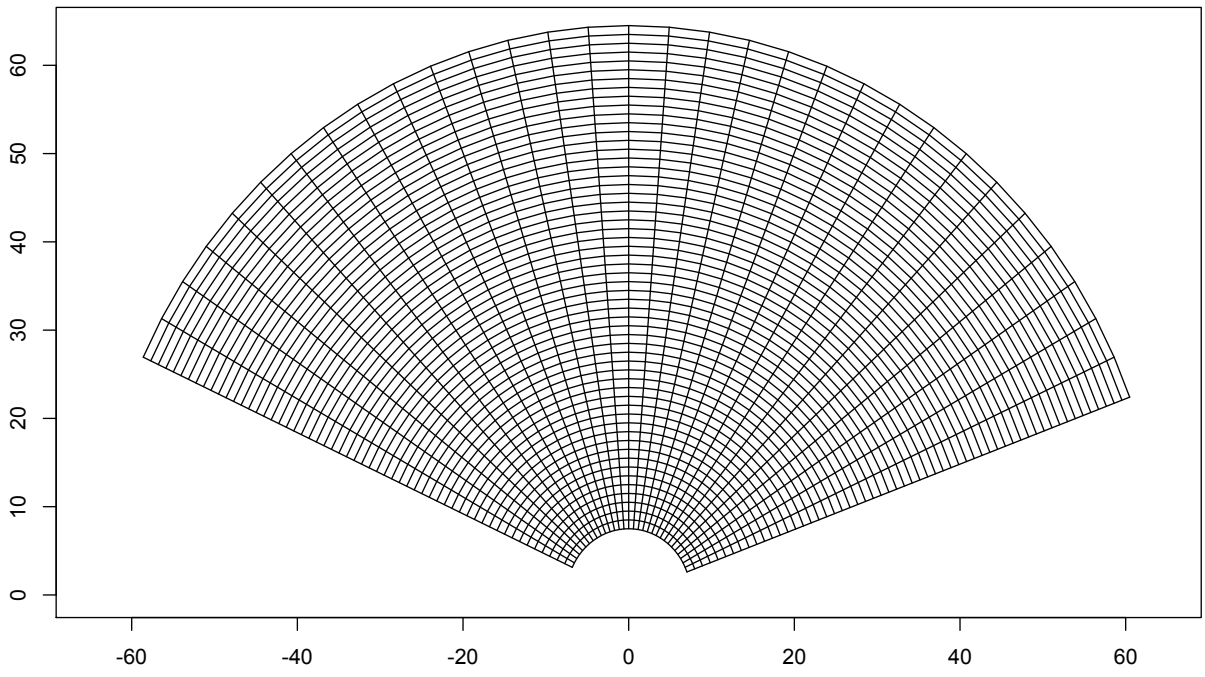


Figure 1: Ultrasound fan consisting of 31 radial scanlines and 64 pixels along each scanline.

References

Kawamoto, A. H., Liu, Q., Mura, K., and Sanchez, A. (2008). Articulatory preparation in the delayed naming task. *Journal of Memory and Language*, 58(2):347 – 365.

McMillan, C. T. and Corley, M. (2010). Cascading influences on the production of speech: Evidence from articulation. *Cognition*, 117(3):243 – 260.

Raeesy, Z., Baghai-Ravary, L., and Coleman, J. (2011). Parametrising degree of articulator

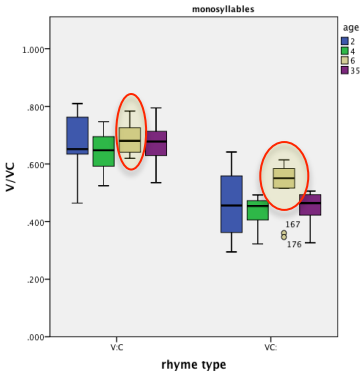
movement from dynamic MRI data. In *12th Interspeech*, pages 2853 – 2856.
Rastle, K., Harrington, J. M., Croot, K. P., and Coltheart, M. (2005). Characterizing the motor execution stage of speech production: Consonantal effects on delayed naming latency and onset duration. *Journal of Experimental Psychology: Human Perception and Performance*, 31(5):1083 – 1095.

Competition for phonetic devices: a cross-linguistic investigation of VC timing

Elinor Payne (University of Oxford)
Brechtje Post (University of Cambridge)
Hanne Simonsen (University of Oslo)
Nina Garmann (Oslo and Akerhaus University College of Applied Sciences)

When learning to speak, infants' exposure to the distributional frequencies (in e.g. syllable structure, etc.) of the ambient language can subtly influence the developmental path followed (Vihman & Velleman, 2000; Prieto et al, 2006) despite common neurophysical constraints as the infant capitalizes on that language's statistical properties to learn its structures (Saffran et al, 2003). When learning to speak, infants have to learn the language-specific structure, and also have to acquire and fine-tune the necessary phonetic skills to be able to map these structures onto speech output. But languages also differ in the way structure is phonetically implemented, making different use of a range of 'phonetic devices', and children also have to learn the correct mapping. A particular phonetic 'device' may also be a multiple signifier, meaning there is interpretational competition for devices. For example, the relative timing of VC sequences is known to be an important cue to coda obstruent voicing in English (e.g. *beat/bead*), but the duration of the vowel in itself may also be an associative cue to vowel quality (e.g. *bit/beat*). Similarly, in Norwegian, VC timing is known to play some role in the cueing of [voice] in coda obstruents, but is also a key marker of syllable weight (e.g. *hat/hatt*). We could say that English and Norwegian present a different ecology of competing demands for the 'same' phonetic device. This paper examines the degree to which different ecologies can impact the acquisition pathway, by examining the acquisition of VC sequences in Norwegian and English, the temporal coordination of which varies systematically in both languages, but in different ways.

We recorded and analysed productions of 9 English-learning and 9 Norwegian-learning children aged 2;6, 4 and 6 years, elicited via a picture-naming task which prompted monosyllabic words containing the relevant features in both languages (e.g. *back, bag, feet, bead, sit, lid*; *bok, smokk, dag, flagg*), and compared these with adult productions. We found that the youngest children already produce adult-like VC relations (as expressed by vocalicness of the rhyme, or V/VC) in both languages, although absolute durations are overly long and precise details as to how timing relations are achieved vary as a result of immaturity in phonetic co-ordination. VC timing relations differentiate [voice] in the coda obstruent by 2;6 years and interact with quantity in Norwegian (with two distinct patterns in English, and four in Norwegian). As phonetic abilities are fine-tuned, and connected speech rate increases, 'mapping' is affected in both languages. In English, we find early use of vowel duration to signal differences between tense/lax vowels, but this shifts to prioritise marking of coda voice by the age of 4 years. In Norwegian, there is an adult-like interaction of [voice] and [quantity] at an early age, but this is undermined by the age of 6, when implementation becomes compromised by the demands of longer domain phonetic co-ordination at a faster speech rate, and the contrast between V:C and VC: is not robustly maintained (see Fig 1, with relative lack of contrast in V/VC for the two rhyme structures at age 6 marked by circles). (We also report an unexpected and hitherto undocumented presence of pre-aspiration of long consonants for Oslo Norwegian, and discuss implications for VC timing relations.)



We conclude that the more crowded ecology of demands for VC timing in Norwegian has a greater impact on Norwegian-learning children at this age, resulting in an apparent U-shaped curve in development (temporary regression from earlier adult-like structures) contrast in.

These findings suggest the acquisition pathway is determined not just by developing articulatory skills and knowledge of a particular phonological system, but also by a play-off between competing uses for duration as a linguistic-

phonetic ‘device’ for implementing that system. In learning to speak, even once children are aware of a phonological structure and have the phonetic skills to implement it, they still have to negotiate the complex mapping between the two. Our study suggests these mappings themselves take some time to fall into place, and competition *between* mappings may be a contributing factor to the time taken and the pathway taken.

Word specific effects in a New Zealand English push chain

Janet B. Pierrehumbert (University of Oxford)
Jennifer B. Hay (University of Canterbury)
Abby J. Walker (Virginia Tech)
Patrick LaShell (University of Canterbury)

It is well known that analogical changes affect rare forms more readily than frequent forms. But what about regular sound changes? On their way to affecting the entire lexicon, do they affect all words at the same time and in the same way? Or are some words leaders, and others laggards, during changes in progress?

Labov (2004) argues that there are no word-specific effects in regular sound change. In contrast, Bybee (2000) suggests that frequent words lead in leniting historical hanges, a suggestion that builds on findings that frequent or predictable words are produced with reduced articulatory effort (c.f. Aylett and Turk, 2004). Labov's claim would follow from a model in which phonetic rules are sensitive only to the phonological form of a word, and not to any other properties. For Bybee's claim to be viable, it is necessary to assume that people have word-specific phonetic memories. Making this assumption, Pierrehumbert (2001, 2002) develops the formal apparatus for frequent words to lead in a sound change by being subjected more often to a persistent leniting phonetic bias. As noted in Kiparsky (2014), the critical prediction is that the difference between frequent and rare words would change once a regular sound change is in progress. A synchronic difference does not, on its own, demonstrate a connection with change over time, since it could arise from a static (time-invariant) difference in articulatory effort.

Here, we summarize a study of the New Zealand Short Front Vowel rotation (Hay et al, 2015) in which the TRAP vowel fronted and raised, colliding with the DRESS vowel, which in turn fronted and raised. In this contrast-preserving push chain, the KIT vowel then lowered and backed. In the Origins of New Zealand English database of speech recordings, this change is delineated over a 137-year span in speaker birthdate (80646 word tokens by 549 speakers). A mixed effects model reveals a significant interaction of word frequency with the place of a word in the shift, supporting the claim that people have word-specific phonetic memories. The effect is small, supporting the claim in Pierrehumbert (2002) that word-specific phonetic patterns take the form of biases within the broader and more abstract phonological system. Surprisingly, it is the low frequency words that lead rather than the high frequency ones. The direction of the effect can be understood by bringing to bear recent experimental results on lexical competition in speech processing. Critical to the explanation is a difference between recognizing a word token, and encoding it in long-term memory.

References

-- Aylett and Turk (2004) *Language and Speech* 47 (1), 31-56.
-- Bybee, J (2000) *Language Variation & Change*, 14, 261–290.
-- Hay, J.B., Pierrehumbert, J.B., Walker, A. J. and LaShell, P. (2015) *Cognition* 139, 83-91.
-- Kiparsky, P. (2014) in *The Routledge Handbook of Historical Linguistics*, Routledge, New York, 64–102
-- Labov, W. (2004) Principles of Linguistic Change, Internal Factors. Wiley-Blackwell, Malden.
-- Pierrehumbert, J.B. (2001) in Frequency Effects and the Emergence of Linguistic Ctructure. John Benjamins, Amsterdam. 137-157.
-- Pierrehumbert, J.B. (2002) in *Laboratory Phonology VII*, Mouton de Gruyter, Berlin, 101–139.

Categorical linguistic and gradient paralinguistic information : ERP evidence

Brechtje Post (University of Cambridge) and Kai Alter (University of Oxford and Newcastle University)

Multiple cues interact to signal multiple functions in intonation simultaneously which makes intonation notoriously complex to analyze. For instance, while some of the intonational information in the signal is interpreted categorically (e.g. a rising pitch contour signals a question instead of a statement in some languages), it can also be more gradient (e.g. a higher rise can sound increasingly excited). Form and meaning are closely intertwined, and the question of how the relationship between them is best formalised in a theoretical model has been a topic of debate for decades.

The main objective of this research is to test the central tenet that underlies the predominant theoretical framework for intonation analysis (the 'Autosegmental-Metrical approach', or AM), according to which categorical, linguistically used information in intonation is crucially distinguished from gradiently varying phonetic information. This theoretical distinction allows us to draw up testable hypotheses about intonational features that are critical in speech comprehension, and to examine the cognitive and neural architecture that supports the processing of the relevant cues in communication. However, evidence for the phonetics/phonology dichotomy on which AM hinges has proved elusive.

Advocating a multidisciplinary approach, we turn to Event-Related Potentials (ERP) in EEG (electroencephalography) to provide direct evidence that different types of intonational information – linguistic/phonological and paralinguistic/phonetic – recruit overlapping but distinct neural systems, which differ not only in their neural architecture, but also in the time-course of activation in the subcomponents of the systems. EEG was recorded while participants performed a categorical perception discrimination task in which categorical and gradient intonational variation were crossed in a 2x2 design. The stimuli were single word utterances (neutral place names) with different pitch contours created by means of resynthesis in Praat: (A) fall of 3 semitones (ST, (B) fall of 9 ST, (C) rise of 3 ST, (D) rise of 9 ST. In line with behavioural and functional Magnetic Resonance Imaging findings obtained for the same stimuli in earlier experiments, the difference in pitch direction (A&B vs. C&D) was expected to be processed primarily as a categorical linguistic distinction (signalling e.g. question vs. statement in the absence of communicative context), whereas the difference in pitch excursion is more likely to be interpreted as a gradiently varying paralinguistic difference (A&B vs. B&D; signalling e.g. different levels of arousal when angry or excited).

A preliminary analysis of the data collected for 15 right-handed native British English participants with no known hearing impairments (analysis of the remaining participants' data is underway) shows that categorical linguistic information elicits a set of positive on-going Event-related potentials (ERPs) which have previously been associated with prosodic processing, including a P300 associated with categorisation more generally, and a P600 which could be interpreted to reflect the processing of 'question' intonation in the absence of syntactical question markers (as in, e.g. '*Hongkong?*'). By contrast, gradient paralinguistic intonation only yields significant ERPs when it coincides with a change in linguistic function.

We argue that the findings can be accounted for in a model in which linguistic (phonological) intonation engages a language-specific fronto-temporal system which is specialised for processing categorical linguistic information, while paralinguistic intonation, which reflects biological imperatives more directly, engages a distributed bilateral system which supports perceptual and cognitive processing more generally.

A phonetic and phonological investigation of the Urdu vowels

Ishrat Rehman, University of Kent, i.rehman@kent.ac.uk
Prof Amalia Arvaniti, University of Kent, a.arvaniti@kent.ac.uk

Literature is sparse for Urdu linguistics, and what is available is often contradictory. There are many disagreements about the phoneme inventory of Urdu and this is particularly evident with respect to the vowel inventory. Most of the literature agrees that Urdu does not have diphthongs, however even that is disputed by the survey and perception experiments of Waqar and Waqar (2003) who argue that Urdu has 13 diphthongs.

Given the disagreement across the literature and the prevalence of cross-language and regional influences, the present study focuses on the vowel system of Urdu as spoken in Punjab, Pakistan and largely uses the inventory proposed by Saleem *et al*. (2002) with slight modifications to address pedagogical goals of the study not discussed here. Twenty-five of the most common vowels found across the available literature were selected for investigation. These included the following: 12 oral vowels (8 long /i/ /e/ /ɛ/ /æ/ /ɑ/ /ɔ/ /o/ /u/ and 4 short /ɪ/ /ʌ/ /ə/ /ʊ/); 7 nasal vowels (/ĩ/ /ẽ/ /ə̃/ /æ̃/ /ɑ̃/ /õ/ /ũ/); 6 diphthongs (/ɑɪ/ /ɔɪ/ /ɑʊ/ /ɪə/ /eə/ /ʊə/). These 25 vowels were embedded in 30 monosyllabic minimal or near minimal pairs. The 5 additional words were used to resolve ambiguity when speakers varied in their pronunciation of target vowels in the standard carrier phrases. The syllable structure for oral and nasal monophthongs was C_1VC_2; for diphthongs the structure was CV. The target vowels were placed in the same environment as much as possible; in the majority of words, the vowels were between a bilabial (/b/or/p/) and an alveolar or dental plosive (/t/, /d/, /ʈ/, or /ɖ/); C_1 was always unaspirated and C_2 was aspirated only in a few cases. The list of words was as follows: /biʈ/ /bɪɖ/ /pet/ /bɛɖ/ /bəɖ/ /bʌɖ/ /bæʈ/ /bɑɖ/ /pɔɖ/ /poɖ/ /bʊɖʰ/ /buɖʰ/ /kuɖ/ /suʈ/; nasal: /bĩʈ/ /pẽʈʰ/ /bə̃ɖ/ /pə̃k/ /sə̃ɖ/ /bæ̃d/ /bɑ̃ɖ/ /sə̃d/ /ʈõɖ/ /bũɖʰ/; diphthongs: /pɑɪ/ /bɔɪ/ /pɑʊ/ /pɪə/ /geə/ /dʊə/.The test words were embedded in two types of sentences: (i) a standard carrier phrase of the form "I will say ___ once" (e.g., /mɛ̃ ɪsɛ **biʈ** ɛk bɑr kəhõ gɪ/); (ii) longer and more varied sentences in which these same words were likely to naturally occur; e.g., /kəi gʰente **biʈ** gəe/ "Many hours have passed".

Twenty-six speakers, 13 males and 13 females from Punjab, Pakistan, took part in the experiment. Most participants were 28-55 years old with two exceptions, a 72 and a 19 year old, both female. All participants were multilingual and spoke at least Punjabi, Urdu, and English. They were from different parts of Punjab and spoke different dialects of Punjabi. They all belonged to elite or upper middle class and were highly educated except for one female who had only GCSE level qualifications from Pakistan. All participants hold British citizenship and have been living in England for 5 to 25 years but use Urdu regularly.

Each participant read from a computer screen five sets of the sentences in pseudo-randomised order such that standard carrier phrases and natural sentences alternated; the Urdu sentences (written in Arabic script in the Nastaleeq style which has an extended Arabic character set) were presented on screen one at a time using PowerPoint. The total number of tokens was 7800 (26 speakers × 30 words carrying the target vowels × 2 sentence types × 5 repetitions).

The test words were recorded in .wav format using a Zoom Handy Recorder H4n with sample settings of 44.1 kHz/16 bit. Acoustic analysis was carried out using Praat. Praat scripts are used to measure the frequencies of the first, second and third formant in three positions: at vowel onset (+10ms), at the middle of the vowel, and at vowel offset (-10ms). These measurements will be used to address questions regarding the vowel inventory of Urdu. In particular they will help to establish the following: whether there is a consistent difference between short and long vowels in Urdu; whether vowels the status of which have been disputed have different qualities; this e.g., applies to the central vowels, /**ə**/, /ʌ/ and /**ɜ**/, front vowels /e/, /ɛ/ and /æ/; and back vowels /o/ and /ɔ/; whether Urdu as spoken in Punjab has nasal vowels, and diphthongs.

References

Fatima, N., Aden, R. (2003). *Vowel Structure of Urdu: CRULP Annual Student Report,* 2002-2003, pp72-77

Saleem, A. M., Kabir, H., Riaz, M. K., Rafique, M. M., Khalid, N., & Shahid, S. R. (2002). *Urdu Consonantal and Vocalic Sounds: CRULP Annual Student Report,* 2001-2002, pp1-5

Waqar, A. & Waqar, S. (2003). *Identification of Diphthongs in Urdu and their Acoustic Properties: CRULP Annual Student Report,* 2002-2003, pp16-23

Phonetic developments in the OED: Spoken Pronunciations and World Englishes

Catherine Sangster, Oxford Dictionaries, OUP
Matthew Moreland, University of East Anglia & OUP Freelancer

This poster will explain and illustrate the strategy and processes involved with two pieces of pronunciation work currently underway on the Oxford English Dictionary. We look forward to sharing our progress with our fellow phoneticians, and hearing their comments and questions.

Every word in the Oxford English Dictionary (3rd edition), unless obsolete, has at least one British and one American pronunciation in the form of an IPA transcription. In December 2015, audio for these words was added to www.oed.com, and more will continue to be added with each quarterly update (see http://blog.oxforddictionaries.com/2015/12/dictionaries-that-talk-oed). OED users can now hear the words aloud, which enhances accessibility and offers an alternative way of understanding the pronunciation. The publication of this audio was the culmination of a major project, which involved taking quarter of a million transcriptions and creating matching sound files for one of the largest dictionaries in the world.

A second connected project is underway to update and expand OED's coverage of a range of World Englishes. Phonemic transcriptions for Australian, New Zealand, South African, Canadian, Caribbean, Scottish and Irish English are being elaborated and modernized. New models are being developed for OED's transcriptions of Singapore and Malaysian, Hong Kong, and Philippine English, with the potential for other varieties to follow. Alongside the work on transcriptions, sound files are being created so that these pronunciations too can be heard aloud.

Perceptual similarity of identical twins across different L1 listeners: the importance of voice quality in Forensic Phonetics

Eugenia San Segundo[1], Paul Foulkes[1,2], Peter French[1,2], Vincent Hughes[1] and Olaf Köster[3]

[1]Department of Language and Linguistic Science, University of York, UK.
[2]J P French Associates, York, UK.
[3]BKA (Federal Criminal Policy Office, Germany).

A round robin test was recently carried out by the *Bundeskriminalamt* (BKA) to evaluate the proficiency of experts performing speaker identification tasks. The test was approached mostly as an auditory evaluation since the technical characteristics of the recordings prevented most experts to carry out specific acoustic analyses. The speakers under comparison were a pair of female German twins. It is well known that these speakers challenge voice recognition because of their similarity (San Segundo 2014). This is true for automatic speaker recognition but also, from a broader phonetic perspective, in identification/recognition tasks performed by both human experts and naïve listeners. Interestingly, the BKA test suggests that the lack of native knowledge of the language spoken by the twins was not a disadvantage for telling these apart. We cannot draw strong conclusions from this test given the idiosyncrasies of the single twin pair used: advanced age and having lived in different dialectal regions. However, informal feedback gathered from the participants suggests that voice quality (VQ) –approached holistically rather than analytically– was the main cue used by non-native listeners (mainly Dutch) to distinguish the twins. This could be due to the fact that the non-native experts lacked formal training in the dialectal differences across Germany. The BKA test calls for the design of a perceptual experiment of a different nature which could shed light on how listeners of different L1s perform when assessing the voice of very similar-sounding speakers. In this study we have tested, with a larger twin sample and under controlled conditions of age and dialect, whether the different L1 of listeners affect the perceptual distances between speakers. Following McDougall (2013), the degree of perceived similarity is measured using Multidimensional Scaling (MDS).

We selected 10 speakers from the corpus collected by San Segundo (2014): 5 pairs of male monozygotic (MZ) twins, all native speakers of Standard Peninsular Spanish (SPS) with no voice or speech pathologies. The original corpus contains 54 speakers (aged 18-54), so for the selection of the 10 speakers of this experiment, some criteria were established. We controlled for similar age (mean: 21, sd: 3.7) and similar mean f0 (mean: 113 Hz; sd: 13 Hz). We also considered the Euclidean distance (ED) between each speaker and his twin in the study by San Segundo & Mompéan (2016). These EDs were based on the perceptual assessment of their VQ using a simplified version of the Vocal Profile Analysis scheme (Laver 1980). The mean EDs between twins in each pair was 0.6 (in Similarity Matching Coefficients these values could range between 0 and 1), which indicates that their VQ was quite similar: on average, at least 6 VQ settings were shared. We set up a Multiple Forced Choice perceptual experiment with 90 different-speaker pairings, i.e. each speaker is compared with everyone else except himself. The stimuli were speech samples of around three seconds, extracted from semi-directed spontaneous conversations. The 90 pairings were presented in random order to listeners, who were asked to indicate the degree of similarity of each stimuli pair on a scale 1 to 5. Two different groups of listeners were recruited to take part in this experiment: 20 native Spanish speakers and 20 native English speakers. The results of this test will potentially reveal the perceptual importance of VQ to distinguish very similar speakers. The implications of these findings will be evaluated in light of the results of the BKA test.

References

San Segundo (2014). *Forensic speaker comparison of Spanish twins and non-twin siblings: A phonetic-acoustic analysis of formant trajectories in vocalic sequences, glottal source*

parameters and cepstral characteristics, Doctoral dissertation, Menéndez Pelayo International University & Spanish National Research Council.

San Segundo & Mompeán (2016) Voice quality similarity based on a simplified version of the Vocal Profile Analysis: A preliminary approach with Spanish speakers including identical twin pairs, *Sociolinguistics Symposium 21*, University of Murcia, Spain, 15-18 June.

Laver, J. (1980) *The Phonetic Description of Voice Quality*. Cambridge: CUP

McDougall, K. (2013). Assessing perceived voice similarity using Multidimensional Scaling for the construction of voice parades, *IJSLL*, Vol. 20 (2): 163-172.

New perspectives on articulatory data

Sonja Schaeffler, James Scobbie and Alan Wrench (Queen Margaret University)

Articulatory techniques have continued to improve over the last decade. For example, there are now high frame rate ultrasound-audio systems capable of capturing aspects of speech production suitable for dynamic research [1], joining Electromagnetic Articulography and Electropalatography, and there is an increased and refined repertoire of articulatory instrumentation to capture phonation, airflow, muscle activation, etc. However, there are still various barriers when it comes to use outside specialised laboratories, let alone everyday use for teaching, or in speech and language therapy. There are also some limits still to the research questions that can be addressed with the type of articulatory data that can be currently obtained.

This paper highlights new means of displaying and capturing articulatory data. Cheap intra-oral cameras (circa £20) or the ubiquitous mobile phone cameras could potentially overcome crucial cost issues and are easy to operate (though safety and hygiene remain concerns). Even these cheap devices can capture articulation - down to the epiglottis and glottis (see Fig. 1) – and this might be enough for some purposes already. For the purpose of addressing different research questions than previously possible, on the other hand, it might be necessary to add new components to existing laboratory setups that afford new views on the speech production process. Such modifications might include e.g. mounting NTSC micro-cameras (and where necessary, LED lights) to the front or side of an ultrasound headset [2] to capture sagittal or coronal views of lip, jaw or even tongue movement (cf. Fig. 2), in addition to ultrasound imaging. To illustrate the use of such enhanced datasets, a selection of recent research will be exemplified: on speech preparation and reaction time, and lip rounding and lip protrusion in Northern Irish and Swedish vowels. Potential applications for different set-ups and devices will be discussed as well as issues that affect all articulatory data collection, i.e. head stabilisation, scaling, video frame rate, choice of speech targets and imageability. We will argue that a mix of new research questions, new teaching and therapy needs as well as the availability of new technical components will help to further develop and open up the field of articulatory research.

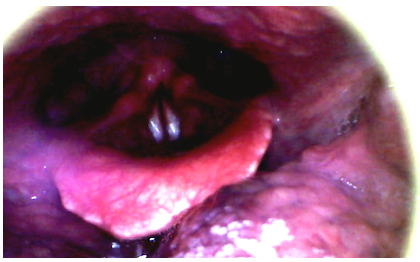


Fig. 1 View with an intra-oral camera.

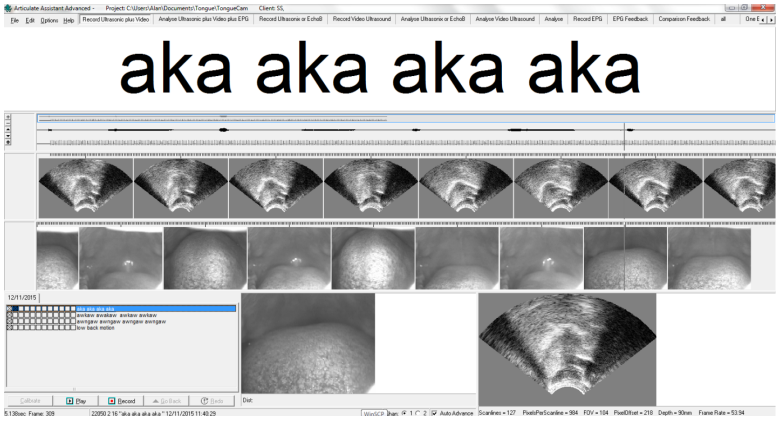


Fig.2 Ultrasound plus coronal video recording of the tongue in AAA.

[1] Wrench, A. and Scobbie, J.M. (2011) Very high frame rate ultrasound tongue imaging. In: Proceedings of the 9th International Seminar on Speech Production, Montreal, pp. 155-162.
[2] Scobbie, J., Wrench, A., & van der Linden, M. 2008. Head-Probe Stabilisation in Ultrasound Tongue Imaging Using a Headset to Permit Natural Head Movement. Proc. 8th ISSP, Strasbourg, 373-376.

Dialectal variation in the interplay of segmental and prosodic timing

Rachel Smith (University of Glasgow) & Tamara Rathcke (University of Kent)

The stressed vowel systems of Standard Scottish English (SSE) and Standard Southern British English (SSBE) differ in terms of quantitative as well as qualitative contrasts. The SSE vowel system lacks several long-short (or tense-lax) oppositions found in SSBE, and the morphological and phonological conditions governing durational allophony differ between the dialects (the Scottish Vowel Length Rule or SVLR, e.g. Scobbie *et al.,* 1999). These durational features of SSE vowels are well established for monosyllables, but poorly understood for polysyllables (Scobbie & Stuart-Smith, 2008). At the same time, duration is involved in signalling of prosodic boundaries and prominence, and these functions can be interdependent (e.g. Byrd & Riggs, 2008; Katsika, 2012). Given the differences in segmental duration between SSBE and SSE, it is intriguing to compare how segmental and prosodic timing interact in the dialects, particularly in words of more than one syllable.

We investigated segmental duration in trochaic disyllables (e.g. *cheesy, picky*) in speakers of SSE from Glasgow and of SSBE from Cambridge (henceforth, "Gla" and "Cam"). We manipulated several factors which can be expected to lengthen vowels:

(1) vowel quality: /i/ vs. /ɪ/, e.g. *peaky* > *picky*

(2) phonological context: following voiced fricative vs. voiceless stop, e.g. *cheesy* > *peaky*

(3) prosodic condition: prosodically prominent vs. non-prominent (nuclear final > postnuclear final) and phrase-final vs. phrase-medial (nuclear final > nuclear non-final):

e.g. *I reckoned it was kind of* ***cheesy*** (nuclear final)

I don't think it's ***remotely*** *cheesy* (post-nuclear final)

That sounds a bit too ***cheesy*** *for them* (nuclear non-final)

Target words were embedded in short scripted dialogues, read by the participant and a trained speaker of their dialect. We present data from three speakers of each dialect (n=432 tokens).

In terms of segmental timing, stressed /i/ was longer in Cam than Gla, as expected, even when followed by a voiced fricative (SVLR-triggering context). Stressed /ɪ/ did not differ between the dialects. The unstressed word-final vowel <-*y*> was longer in Gla than Cam, especially when the stressed vowel was /i/. Consonant durations did not differ between the dialects, except that Gla had significantly more perseverative voicing in medial consonants (*peaky, cheesy)*.

Prosodic timing patterns were found to be affected both by segmental identity, and by dialect. Overall, short stressed vowels (i.e., /ɪ/ in both dialects, and /i/ in Gla) showed less prosodic lengthening than long vowels. Phrase-final lengthening occurred in both dialects, extending backwards from the phrase boundary to the onset of the target word's stressed syllable. However, accentual lengthening (AL) differed dramatically across the dialects: while in Cam we found robust evidence of AL throughout the word, in Gla AL only affected the initial consonant; there was no accentual lengthening of Gla stressed vowels, and final <-*y*> underwent accentual shortening. We consider a number of explanations for this result, including the need for vowels to maintain distinctive duration, and (in an Articulatory Phonology framework) differences in coupling relations among prosodic gestures. Results may inform the development of multi-dimensional models of speech timing.

References

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

Katsika, A. (2012), *Coordination of prosodic gestures at boundaries in Greek*, PhD thesis, Yale University.

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

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

Voice Onset Time in Serbian-English Bilingual Children

Mirjana Sokolović-Perović
University of Reading and University of Warwick

The acquisition of the voicing contrast in bilingual children has been a topic in phonetic research for many years. Previous research has shown that young bilingual children have difficulty acquiring two different VOT systems in languages such as German and Spanish (Kehoe et al., 2004), English and Spanish (Deuchar & Clark, 1996; Fabiano-Smith & Bunta, 2012), and English and Japanese (Johnson & Wilson, 2002). For older children, Khattab's (2000) study of Arabic-English bilinguals found that some children were able to produce short lag and long lag VOT categories comparable to their monolingual peers in both Arabic and English. Netelenbos et al. (2015) found that although English-French bilingual children were able to differentiate between French /p, t, k/ and English /p, t, k/, they employed non-native categories of lag VOTs. The lead VOT category, however, seems to be problematic for Arabic-English and English-French bilinguals. Both in Arabic and French the children produced short lag VOTs instead, but Arabic-English bilinguals used different VOT values in Arabic and English (Khattab, 2000), while English-French bilinguals used only one phonetic VOT category in both languages (Netelenbos et al., 2015).

In the present paper, the realisation of the voicing contrast is examined in a group of Serbian-English bilingual children. Serbian, like French, Spanish and Arabic, has a two-way voicing distinction between lead and short lag plosives, which is different from the English voicing contrast. The study examines VOT production in children aged 7 to 11 years in Serbian and English, and in comparison to their monolingual peers. The findings are discussed in the light of previous research on bilingual VOT acquisition, and in relation to the input they received and the patterns of language use.

References

Deuchar, M. & Clark, A. (1996). Early bilingual acquisition of the voicing contrast in English and Spanish, *Journal of Phonetics*, 24, 351-365.

Fabiano-Smith, L. & Bunta, F. (2012). Voice onset time of voiceless bilabial and velar stops in 3-year-old bilingual children and their age-matched monolingual peers. *Clinical Linguistics and Phonetics*, 26 (2), 148-163.

Johnson, C. & Wilson, I. (2002). Phonetic evidence for early language differentiation: Research issues and some preliminary data. *International Journal of Bilingualism*, 6, 271-289.

Kehoe, M., Lleo, C. & Rakow, M. (2004). Voice onset time in bilingual German-Spanish children. *Bilingualism: Language and Cognition*, 7, 71-88.

Khattab, G. (2000). VOT in English and Arabic bilingual and monolingual children, *Leeds Working Papers in Linguistics and Phonetics*, 8, 95-122.

Netelenbos, N., Li, F. & Rosen, N. (2015). Stop consonant production of French immersion students in Western Canada: A study of voice onset time. *International Journal of Bilingualism*, 1-12.

F2 AND ARTICULATORY FRONTING - PERSPECTIVES FROM HIGH-BACK VOWELS IN SBE

Patrycja Strycharczuk & James M. Scobbie
CASL, Queen Margaret University

The value of the second formant is commonly interpreted as a correlate of degree of articulatory fronting, although some limitations to this articulatory metaphor are widely acknowledged in case of rounded vowels, such as English /uː/ and /ʊ/, since lip rounding affects the F2. In this paper we discuss to what extent F2 correlates with the tongue position and horizontal displacement in the production of /uː/ and /ʊ/ in two types of consonantal context.

Our data are time-synchronised high-speed ultrasound and acoustic recordings of /uː/ and /ʊ/ in monosyllabic words ending in a final coronal consonant (e.g. *food*, *foot*), or in a final liquid (e.g. *fool*, *full*). The words were embedded in a standard carrier phrase, and read by five younger (under 25) and five older (over 45) female speakers of Southern British English. For each speaker, we imaged the occlusal plane, by asking them to bite on a piece of plastic (a bite plate), and press their tongue against it. We then measured the distance between the point where the back of the tongue crosses the occlusal plane and the x-value corresponding to the position of the highest point of the tongue during an average /i/ pronunciation (Figure 1). This measure was extracted at 10% intervals throughout the vowel or vowel+liquid sentence. The distances were z-score normalised to reduce the effect of anatomical differences between speakers, and compared to the F2 measurements at corresponding time points. The F2 measurements were normalised by subtracting the average F2 value for /i/.

Figure 1

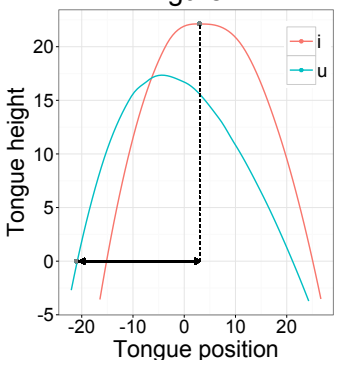
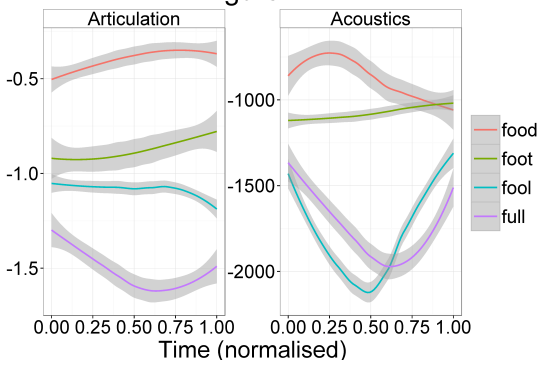


Figure 2 compares the two types of analysis for an example speaker, YF1. The average differences between *foot* and *food* and their dynamic profiles are similar in articulation and acoustics. The vowel+liquid sequences, however, have lower F2 than could be inferred from the degree of articulatory backness. In addition, their relative position is different in articulation and acoustics: the articulatorily more retracted *full* may have higher F2 values than *fool*. Similar generalisations were obtained in the statistical analysis of the aggregated data. We fitted an additive model predicting the tongue position depending on the normalised F2 and its interactions with context and time point. In general, F2 and tongue backness are correlated for *food* and *foot*, somewhat correlated for *full*, and largely uncorrelated for *fool*. We discuss these findings using illustrations from entire tongue contours, to gauge the more holistic effect of tongue posture on acoustic resonances. We further consider apparent time differences in the degree of articulatory and acoustic /uː/- and /ʊ/-fronting in different consonantal contexts, based on additive models including interactions with speaker age.

Figure 2



Is gender a factor in real-time acoustic change in coda liquids in Glasgow?

Jane Stuart-Smith, Robert Lennon, Rachel Macdonald, Farhana Alam*, Ludger Evers^
Duncan Robertson, Márton Sóskuthy+; Eleanor Lawson~

English Language/GULP, Mathematics and Statistics^, University of Glasgow;*
Language and Linguistic Science, University of York+; Queen Margaret University Edinburgh~

Variationist sociophonetic descriptions of sound change in progress have often shown that female speakers lead change, though this can depend on local regional and sociolinguistic context (Labov 2001; Eckert 2000). There is relatively less evidence for the role of gender in real-time change at least partly given that real-time studies are comparatively less common. Vernacular Scottish English, especially that of Scotland's largest city, Glasgow, seems to have experienced a number of changes over the course of the 20th century (Stuart-Smith et al to appear). To date, previous real-time studies of this accent have not considered the role of gender in sound change.

This paper focuses on the coda liquids /l r/ in e.g. *tell, tear*. Previous impressionistic phonetic analysis of coda /r/ in Central Belt Scottish English has suggested weakening of articulated /r/ especially in vernacular speakers; the pattern according to gender is less clear (e.g. Lawson et al 2014). A previous dynamic acoustic study of coda /r l/ in 8 male Glaswegian vernacular speakers from four decades showed differences in F3 for /r/ and /F2/ for /l/ related to decade of birth consistent with weakening of /r/ and darkening of /l/ over time (Stuart-Smith et al 2015). There is no acoustic phonetic study of this dialect which considers gender, either for an apparent-time, or a real-time investigation, for either of the coda liquids.

This paper analyses the speech of 24 speakers of Glasgow vernacular, half male and half female, aged between 67-90 years old, with six recorded talking in oral history interviews from the 1970s, 1980s, 1990s, and 2000s. The first 35 usable tokens of coda /r l/ were extracted from a force-aligned electronic spoken corpus, coding for preceding vowel quality, following segmental context (e.g. *tell and*), decade of birth and gender. Tokens followed by another liquid were excluded (e.g. *tell like*), as were forms showing historical Scots L-vocalization, e.g. *a'* for *all*. Following Plug and Ogden 2003, we carried out a parametric analysis of the rhyme (vowel+liquid sequence), and so manually labelled the phase from the onset of periodicity for the vowel to the last discernible cycle of the liquid/end of the segment if the liquid was very weak. Formant tracks for the first three formants were obtained using standard settings in Praat, and manually corrected using Formant Editor. An additional auditory analysis of /r/ was carried out using a randomized Multiple Force Choice test in Praat, in order to assess the extent to which dynamic formant analysis of /r/ might be affected by spectral discontinuity arising from tapped and the very few trilled /r/s.

Possible differences in the realization of the liquids across the course of the vowel+liquid sequence with respect to phonetic and social factors will be assessed using functional data analysis applied to the formant tracks. The results will be presented according to preceding and following segmental context, decade of birth, and gender, in order to uncover the role of gender in real-time change in coda liquids in this variety of Scottish vernacular English. We will also consider the efficacy of using formant tracking over the rhyme to achieve a general characterization of coda /r/ across a range of auditorily defined articulatory variants.

References

Eckert, P. 2000. *Linguistic Variation as Social Practice,* Oxford: Blackwell.
Lawson, E., Scobbie, J. and Stuart-Smith, J. (2014), 'A socio-articulatory study of Scottish rhoticity', in R. Lawson (ed.), *Sociolinguistics in Scotland,* Basingstoke: Palgrave Macmillan, 53-78.

Labov, W. 2001. Principles of Linguistic Change, Vol. 2. Oxford: Blackwell.
Plug, L. and Ogden, R. 2003. A parametric approach to the phonetics of postvocalic /r/ in Dutch. *Phonetica* 60. 159–86.
Stuart-Smith, J., Jose, B., Rathcke, T., Macdonald, R. and Lawson, E. (to appear), 'Changing sounds in a changing city', In E. Moore and C. Montgomery (eds), *A Sense of Place,* Cambridge: Cambridge University Press.
Stuart-Smith, J., Macdonald, R., José, B., Robertson, D., Sóskuthy, M., Evers, L. (2015), 'A dynamic acoustic view of real-time change in word-final liquids in spontaneous Glaswegian', *Proceedings of the 18th International Congress of Phonetic Sciences*, 10-14 August 2015, Glasgow.

Studies of Obstruent Devoicing: Methodological Implications

Marianne Thurgood, University of Sheffield

Obstruent devoicing is a well-studied phenomenon. Sibilants' tendency to be realised as devoiced has been studied considerably more than that of non-sibilant fricatives (Smith 1997) and several studies have shown that female speakers devoice fricatives more frequently than males (e.g. Verhoeven et al 2011). The present study examined /v/ in English. Its aim was to determine whether a fricative with less strong fricative noise, i.e. a non-sibilant, exhibited similar patterns of devoicing both in different word positions and by speakers of different sexes i.e. more commonly devoiced by females and in word-final/pre-pausal position. Using 'The Rainbow Passage' read aloud by eleven male and seventeen female participants in the UCL Speaker Database (Markham & Hazan 2002) two word-initial, two word-medial and one word-final token of /v/ in intervocalic position as well as one pre-pausal token were obtained, giving a total of 168 tokens. Using the electroglottographic signals included in the corpus, each token was analysed to determine what percentage of its duration, if any, was produced without vocal fold vibration.

Researching the phenomenon uncovered that the results of some devoicing studies vary. Smith (1997) finds that women devoice 56% of tokens of /z/, while Verhoeven et al (2011) find that women devoice 72% of fricative tokens. This is most likely because the boundary between what is classed as 'devoiced' and what is classed as 'partially devoiced' has been set differently by different researchers, meaning that the term 'devoiced' used in one study does not mean the same thing as 'devoiced' in another (see Table 1).

Table 1: Categorical devoicing boundaries used in previous studies

	Devoiced	**Partially devoiced**	**Fully voiced**
Smith (1997)	0 - 25%	25 - 90%	90 - 100%
Jesus & Shadle (2002)	0 - 33.3%	33.3 - 50%	50 - 100%
Verhoeven et al (2011)	0 - 66.6%	N/A	66.6 - 100%

Although each study achieves its own objective in calculating how often fricatives are devoiced, a meta-study of devoicing highlights several problems with these categories:

- they are inconsistent and seemingly arbitrary
- a slight adjustment of the boundaries can make many more tokens fall into a certain category, causing a particular phenomenon to appear more common than it is, or at least more common than it was found to be in other studies with different (lower) boundaries
- two tokens with a very similar pattern of devoicing can be classed in different categories if they fall just either side of a boundary

I propose a gradient approach, whereby the tokens are not divided into categories at all, but a mean of their 'percentage devoiced' figures ranging from 0-100% is calculated. The above studies' methodologies involve calculating a percentage of the segment's duration which is devoiced, and going on to group these into binary or ternary categories obscures the gradience of devoicing as a phonetic (as opposed to phonological) process. A sample of my results for the fricative /v/ in the three word positions and pre-pausal position are presented in Table 2. In line with previous results, women were found to devoice not only more often but *for longer* than men, however these results show the phenomenon is not as prevalent as categorical studies might suggest. This is a way of analysing the data that does not impose categories on a gradient phenomenon.

Table 2: Results

Mean 'percentage devoiced' figures			% of tokens produced with 0% voicing
a **v**ery	Women	40%	18%
	Men	19%	0%
divisio n	Women	45%	18%
	Men	17%	0%
ha**v**e accepte d	Women	23%	6%
	Men	12%	0%
abo**v**e	Women	95%	53%
	Men	92%	45%

References

Jesus, L. M. T. & Shadle, C. H. (2002) A parametric study of the spectral characteristics of European Portuguese fricatives. *Journal of Phonetics* 30(1), pp.437-464.

Markham, D. & Hazan, V. (2002) UCL Speaker Database. *Speech, Hearing and Language: Work in Progress* 14(1), pp.1-17.

Smith, C. L. (1997) The devoicing of /z/ in American English. *Journal of Phonetics* 25(4), pp.471-500.

Verhoeven, J., Hirson, A. & Basavaraj, K. (2011) The devoicing of fricatives in Southern British English. *Proceedings of the 17th International Congress of Phonetic Sciences,* pp.2070-2073. Available at: http://www.researchgate.net/publication/228837996_The_Devoicing_of_Fricatives_in_Southern_British_English [Accessed 29 Oct 2015].

Phonetic details and the projection of more talk

Gareth Walker
University of Sheffield

Previous research in phonetics has provided insights into how speakers signal that their talk is coming to an end and that a conversational partner may speak (see e.g. Bögels & Torreira, 2015; Caspers, 2003; Ogden, 2001; Schaffer, 1983). Comparatively less is known about how speakers manipulate the phonetic details of their speech to indicate that they have more to say. This paper builds on Local and Walker (2012) to offer a systematic study of what they identified as talk-projecting features: phonetic features which indicate that the current speaker will continue talking. It does so by exploring in detail what happens at points of possible syntactic completion i.e. points at which a speaker may, phonetic design notwithstanding, be interpreted as having finished what they have to say.

The data are taken from the CALLHOME corpus (Canavan, Graff, & George, 1997) which is made up of dual-channel recordings of spontaneous telephone conversations. A total of 12 calls were selected for analysis. Selections were made to balance numbers and combinations of male and female callers and receivers of calls. The focus here is on the callers' speech in the transcribed portions of the corpus. Callers were selected to cover a range of ages (the age range of the selected callers is 18-54 yrs; mean = 32 yrs). The callers grew up in several different US states.

For each call, the first 20 transcribed points of possible completion in the callers' speech which were followed by same-speaker talk were identified. All 240 points were then subjected to auditory and computer-based phonetic analysis. The nature and distribution of the following talk-projecting features is discussed: avoidance of lengthening, articulatory anticipation, continuation of voicing and reduction of consonants and vowels. Of the points of possible syntactic completion followed by same-speaker talk 183 (76.3%) were found to have one or more of the talk-projecting features identified above. Conversational partners usually avoided talking in response to points of possible syntactic completion where talk-projecting features occurred (76.6% of cases). Somewhat surprisingly, talk-projecting features were evident in a similar proportion of cases where conversational partners began to speak (19 out of 26 cases; 73.1%). Analysis of the interactions reveal that any incoming speech following a point of possible completion accompanied by talk-projecting features is constrained in various ways.

References

Bögels, S. & Torreira, F. (2015). Listeners use intonational phrase boundaries to project turn ends in spoken interaction. Journal of Phonetics, 52, 46–57.
Canavan, A., Graff, D., & George, Z. (1997). CALLHOME American English Speech LDC97S42. DVD. Philadelphia, PA: Linguistic Data Consortium.
Caspers, J. (2003). Local speech melody as a limiting factor in the turn-taking system in Dutch. Journal of Phonetics, 31, 251–276.
Local, J. & Walker, G. (2012). How phonetic features project more talk. Journal of the International Phonetic Association, 42(3), 255–280.
Ogden, R. (2001). Turn transition, creak and glottal stop in Finnish talk-in-interaction. Journal of the International Phonetic Association, 31(1), 139–152.
Schaffer, D. (1983). The role of intonation as a cue to turn taking in conversation. Journal of Phonetics, 11(3), 243–257.

The role of rhythmic attending in improving connected speech comprehension for learners of English as a Foreign Language

Ewa Wanat (University of Glasgow)

In casually spoken English, reduction processes can dramatically affect the phonetic shape of words, especially function words, reducing their intelligibility for non-native listeners. There is a close connection between reduction and speech rhythm: metrically weak syllables reduce more, and may be cued only by subtle phonetic detail that non-native listeners struggle to detect. Despite growing evidence that attention to speech and music is rhythmically guided (e.g., [1] [2]) and that speech processing depends on language rhythm (see overview in [3]), little work has tested whether encouraging non-native learners to attend to rhythm might support their comprehension of casual speech. We report an experiment to test whether learners' casual speech comprehension is affected by the rhythmic organisation of speech they are exposed to.

Participants (n=62) were pre-intermediate to upper-intermediate learners of English, resident in Glasgow. The experiment had three phases, pre-test, exposure, and post-test. In all phases, learners heard short sentences spoken fast and casually by native speakers of Glasgow English (different sentences in each phase). In the pre- and post-test learners performed a transcription task involving filling in gaps corresponding to the sentences' function words, e.g. the answer to ___ *stuck* ___ *seat* ___ *car* would be *It is stuck to the seat in the car*. In the exposure phase, learners listened to sentences produced several times with increasing speech rate. In the rhythmic condition, the sentences had regular metrical structure (e.g. *He was **fat** for a **cat** in a **box***, stressed syllables in bold) and tokens were elicited through a modified version of the speech cycling procedure (Cummins and Port, 1998) in which speakers align the stressed syllables with metronome beats occurring at gradually increasing rates. In the non-rhythmic condition, sentences had irregular metrical structure (e.g. *It was **big** even for a **chick**en on a **farm***) and tokens were elicited by instructing speakers to read a story at different rates. We predicted that hearing materials with greater rhythmicity would lead to greater improvement from pre- to post-test.

A mixed-effects logistic regression analysis showed a trend for learners' comprehension of function words to improve more from pre- to post-test if they received rhythmic rather than non-rhythmic exposure, as predicted. Significant effects of the learners' English language proficiency and the type of connected speech phenomenon were also found: performance improved from pre- to post-test on some of the function word categories, but remained stable or even dropped for others. Ongoing work investigates the detailed relationships between rhythmic structure, grammatical category, and test improvement, and seeks to relate the perceptual results to acoustic properties of the stimuli.

References:

[1] Large, E. W. (2008). Resonating to musical rhythm: Theory and experiment. In S. Grondin (Ed.) The Psychology of Time. West Yorkshire: Emerald.

[2] Dilley, L. C., & McAuley, J. D. (2008). Distal prosodic context affects word segmentation and lexical processing. *Journal of Memory and Language*, 59, 294-311.

[3] Cutler, A. (2012). Native Listening: Language Experience and the Recognition of Spoken Words. Boston: MIT Press.

[4] Cummins, F., & Port, R. F. (1998). Rhythmic constraints on stress timing in English. *Journal of Phonetics, 26*(2), 145–171.

Voiceless Mandarin Chinese vowels

Shichao Wang (Dalian Institute of Technology) and Yang Li (University of Cambridge)

While vowel devoicing has been extensively studied (see Gordon, 1998), there are few acoustic studies on Mandarin Chinese or on Chinese languages. Duanmu (2002: 300–302) offers examples of Mandarin disyllabic words with habitually devoiced vowels, and point to high vowels, aspiration in the preceding obstruent, de-stressing (creating "neutral tone") and low tone as environments conducive for devoicing. Early observations from Stanford Phonology Archive (Crothers, 1979), reported in Gordon (1998), suggest that Mandarin vowel devoicing occurs exclusively in high vowels, and is more likely following aspirated affricate and fricatives. Relatedly, data from one Shanghai Wu Chinese speaker (Zee, 1990) show a tendency for low-toned, word-final and higher vowels to devoice.

An experimental study of Mandarin Chinese vowel devoicing seems opportune, given our limited understanding (Gordon, 1998) of pitch conditioning in devoicing: cross-linguistically, devoicing is inhibited by high tones, when created by pitch accent (Japanese, Nielsen, 2015), question intonation (Greek, Kaimaki, 2015) and lexical high tone (Acoma, Miller, 1965). The present study therefore examines the realisations of Mandarin Chinese vowels by varying tonal, segmental and intonational conditions.

I report a production experiment of 20 Mandarin Chinese speakers located in the city of Dalian (10M, 10F; mean age=22, SD=3.2). So far, data from 10 speakers have been analysed. The stimuli consist of disyllabic words which vary systematically in tones and segmental composition of the second syllable. All four lexical tones, in addition to "neutral tones", are included. Syllable types include [fu], [su], [ɕi], [tɕʰy] and [tʰa]. These are embedded in the final position of two segmentally similar carrier sentences, one statement (/wo ɕiaŋ ʂʷo tə tsʰz ʂʐ̩XX/. "The word I want to say is XX".), one yes-no question (/ni ɕiaŋ ʂʷo tə tsʰ ʂʐ̩XX/. "Is XX the word you want to say?"), where a boundary H tone is observed (Peng et al., 2005).

The duration of voicing and duration of the preceding consonant are extracted. These are tested for correlation to establish a possible trading relationship (Eftychiou, 2007; Oberly & Kharlamov, 2015). Vowel tokens were also divided into the following categories: a) non-devoiced, b) completely elided and fricated, c) partially devoiced, when the voicing lasts less than 30 ms (Jun, Beckman, Niimi, & Tiede, 1997). Hypothesis testing was performed with multinomial logistic regression.

Several general trends can be observed from the experimental data. First, devoicing is strongly inhibited (<1%) when the vowel carries a high tone or rising tone (T55/T35) or when situated at the end of a yes-no question. On the other hand, complete and partial devoicing are likely (65%) on vowels with neutral tone, followed by low tone (27%) and falling tone (7%). Second, while the implicational hierarchy of more devoicing in high vowels is respected, devoicing can occur on low vowels (8%), contrary to Gordon's (1998) observation. Third, there is significant negative correlation between the duration of voicing and the duration of onset consonants. Finally, it is clear that, for a subset of speakers, devoicing has been lexicalised on certain high-frequency words which carry final neutral tone (e.g. /tʰuŋ.ɕi/ "things"; /kao.su/"tell") such that the incidence of devoicing and frication is almost at a ceiling level.

One hypothesis to emerge regarding the pitch-devoicing relation is that low pitch environments may create adverse perceptual condition for recovering voicing. Notably, voicelessness tends to co-occur with shorter duration (neutral tone), lower intensity and non-modal phonation, especially creaky voice. In low tones and in the second half of falling tones (a region critical for perception), devoicing seems to be complementarily distributed with creaky voice, whose irregular pulses may create similar percept to voicelessness. By contrast, rising intonation and high/rising lexical tones do not present similar recoverability problems. I argue that perceptual consideration should figure into an eventual account of why high tones inhibit voicing, which is usually explained in articulatory terms (Gordon, 1998; Kaimaki, 2015).

When /i/ is not a high front vowel: An acoustic and articulatory study of Swedish Viby-i

Fabienne Westerberg, University of Glasgow & Queen Margaret University Edinburgh
f.westerberg.1@research.gla.ac.uk

The high front vowel /i/ is often used as an anchor point for phonetic analysis, e.g. to contextualise other vowels, or to normalise acoustic data. But what happens when /i/ is not a high front vowel? This is the case in some varieties of Swedish, where /i/ is not realised [i], but instead has an unusual "thick", "dark" or "damped" quality which appears to be centralised in the acoustic space [1]. This vowel sound is commonly known as *Viby-i* and can be found in "several scattered dialects [across Sweden], both in rural areas and in the city dialects of Stockholm and [Gothenburg]" [2]. Acoustically, Viby-i is characterised by a low F2 and high F1 compared to [i] [1], but its articulation is long disputed as there seem to be several ways of producing this vowel, and articulatory data is scarce. In my paper I will present results from a recent acoustic and articulatory study of Viby-i, and discuss some of the challenges of working with a vowel that does not conform to traditional definitions.

The study uses word list data from 13 Swedish speakers with Viby-i in their speech. The position of /i/ in the vowel space is contextualised by /e ɑ u ʉ/. The acoustic properties of these vowels will be discussed in terms of F1 and F2, with audio examples. Their articulation will be illustrated using data from ultrasound tongue imaging (UTI), a technique which creates a two-dimensional video image of the tongue surface when an ultrasound probe is held under the chin. Mean tongue contours are then derived using bespoke software. Articulatory video of the lips will also be presented.

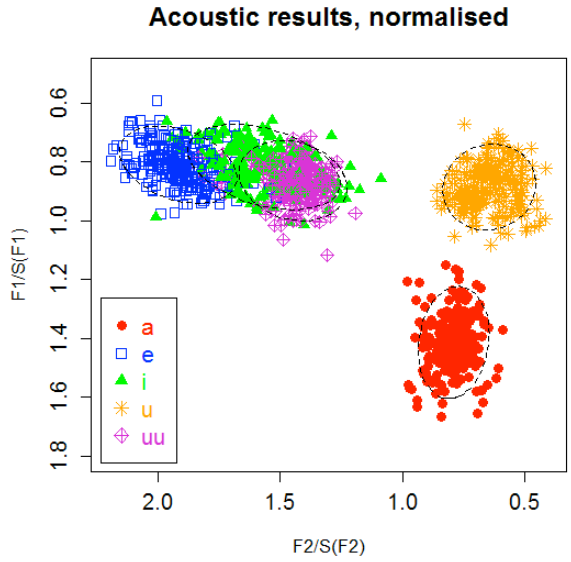


Fig. 1: F1/F2 plot of all vowel tokens. a =/ɑ/, uu =/ʉ/.

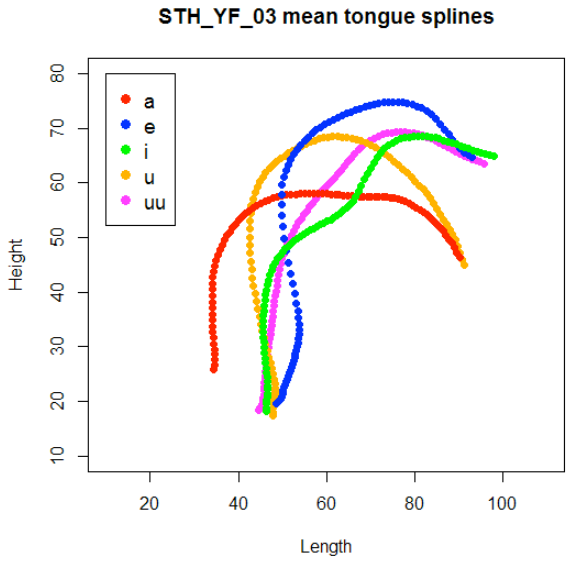


Fig. 2: Profile tongue contours of one speaker, facing right.

The acoustic results confirm that Viby-i is centralised in the acoustic space, having a low F2 and high F1 (Fig. 1), but the articulatory results are more complex. Viby-i is produced with a low tongue body, but the highest point of the tongue is unexpectedly fronted, and the back of the tongue is retracted. Many speakers also use an unusual "double-bunched" tongue shape (Fig. 2). The lips are spread but lax, and the tongue tip can often be seen resting against the lower teeth. There is much variation in the sample, but stratification was only found in an acoustic difference between urban and rural speakers.

The results raise several questions, not least about the theoretical framework of vowel analysis. Can we trust acoustic data to provide a reliable image of vowel articulation? Does the vowel quadrilateral allow us to express sufficient information about vowels? And are vowels generally prone to this degree of individual variation, or could there be fine-grained systematic variation at work? These questions continue to be examined in an ongoing project about Viby-i.

[1] Engstrand, O., Björsten, S., Lindblom, B., Bruce, G. & Eriksson, A. (1998) 'Hur udda är Viby-i? Experimentella och typologiska observationer'. *Folkmålsstudier* 39, 83-95.
[2] Björsten, S. & Engstrand, O. (1999) 'Swedish “damped” /i/ and /y/: Experimental and typological observations'. *Proceedings of the 14th ICPhS*, San Francisco, 1,957-1,960.

Predictive perceptual mechanisms and the interpretation of timing cues to prosodic structure

Laurence White
Plymouth University

Variation in the pitch, length and loudness of speech sounds informs listeners about speech structure. Whilst languages differ in how such prosodic features are organised, some form-function associations may be universal. Firstly, the valency of interpretation of prosodic features appears consistent: sounds that are longer, louder or higher in pitch are more salient. Furthermore, the "Iambic-Trochaic Law" captures the observation that sounds made salient through higher pitch or greater loudness are interpreted as sequence-initial, whilst lengthened segments are interpreted as sequence-final (e.g., Hays & Diehl, 2007).

In accordance with the Iambic-Trochaic Law, phrase-final lengthening of vocalic nuclei and following coda consonants is ubiquitous and has been shown to be a cue to an upcoming boundary (e.g., Price, Ostendorf, Shattuck-Hufnagel & Fong, 1991). However, onset consonants are also found to be lengthened word- and phrase-initially in several prosodically diverse languages (e.g., Cho & Keating, 2001; Cho, McQueen & Cox, 2007). Could segmental lengthening be an effective cue for listeners to both preceding and following boundaries?

In a series of artificial language learning experiments, with native speakers of English, Hungarian and Italian, we explore the importance of the localisation of such timing cues to speech structure, and find cross-linguistic support for an apparent functional division between vowel lengthening and consonant lengthening. In particular, native speakers of all three languages interpret consonant lengthening as indicative of an immediately preceding boundary, contrary to the standard Iambic-Trochaic Law analysis. Attempting to reconcile these diverse findings, we suggest that the interpretation of such timing events may be mediated by a predictive perceptual mechanism tracking the occurrence of successive perceptual centres in the ongoing stream of spoken syllables.

References

Cho, T., Keating, P.A., 2001. Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics*, *29*, 155-190.

Cho, T., McQueen, J., & Cox, E. (2007). Prosodically driven detail in speech processing: The case of domain-initial strengthening in English, *Journal of Phonetics*, *35*, 210-243.

Hay, J.S., & Diehl, R.L. (2007). Perception of rhythmic grouping: Testing the iambic/trochaic law. *Perception and Psychophysics*, *69*, 113-122.

Price, P. J., Ostendorf, M., Shattuck-Hufnagel, S., and Fong, C. (1991). The use of prosody in syntactic disambiguation, *Journal of the Acoustical Society of America*, *90*, 2956-2970.

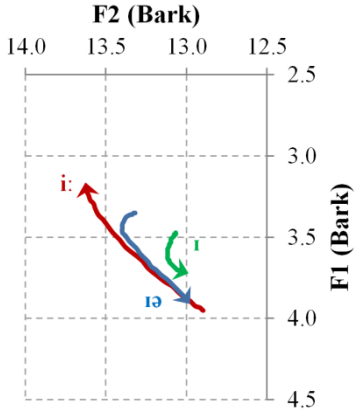
Bead, *bid* or *beard*? Perceiving linguistic and indexical information in Australian English vowels

Daniel Williams
Area of Excellence – Cognitive Sciences, Linguistics Department, University of Potsdam MARCS Institute for Brain, Behaviour and Development, Western Sydney University
ARC Centre of Excellence for the Dynamics of Language

English vowels can generally be well separated with reference to vowel inherent spectral change (VISC). As well as aiding perceptual identification (Hillenbrand, 2013), VISC can convey indexical information, e.g., regional/generational background (Jacewicz & Fox, 2012), and information for forensic voice comparison (Morrison, 2013).

The present study investigates the influence of VISC on listeners' categorisation of Australian English (AusE) front vowels into /iː, ɪ, ɪə/ and its effect on likely speaker characteristics. Elvin *et al.* (under review) found AusE /iː, ɪ, ɪə/ as produced by Western Sydney (WS) speakers are virtually identical in terms of mean first and second formants (F1 and F2) and differ in duration (/iː/ and /ɪə/ are >1.3 times longer than /ɪ/) and in magnitudes and directions of F1 and F2 change (Figure 1). Furthermore, Harrington *et al.* (1997) found distinct patterns of realisation for these (and other) AusE vowels that correlated with, e.g., generation, gender and socio-stylistic accent.

Figure 1. Average trajectories of /iː-ɪ-ɪə/ by AusE males from Elvin *et al.* (under review)



Using electroencephalography, Experiment 1 investigated WS AusE listeners' mismatch negativity (MMN) responses arising from a change in vowel stimulus. VISC of the stimuli in two blocks corresponded to a change in vowel category, while in two further blocks could correspond to between-speaker variation. In the first behavioural task of Experiment 2, WS AusE listeners heard a vowel stimulus from a continuum in which VISC and duration varied in quasi-equal steps and assigned a vowel label (one of /iː, ɪ, ɪə/). In a second task, listeners heard a subset of the same stimuli, which this time also varied in fundamental frequency (f0), and selected likely speaker characteristics.

MMNs from Experiment 1 show listeners were sensitive to changes in the stimuli, suggesting the selected acoustic parameters are used at a pre-attentive level. Experiment 2 shows that listeners' vowel categorisation varied as a function of VISC and duration in line with the three vowels' acoustic properties in speech production. As for speaker characteristics, responses were consistent with respect to f0 but diverse with respect to VISC, suggesting individual interpretations of within-vowel category VISC; however, the likelihood of a speaker as being Australian correlated with VISC across listeners.

References

Elvin, J., Williams, D. and Escudero, P. (Under review). "Dynamic acoustic properties of the monophthongs and diphthongs of Western Sydney Australian English", *Journal of the Acoustical Society of America.*

Harrington, J., Cox, F. and Evans, Z. (1997). "An acoustic phonetic study of broad, general, and cultivated Australian English vowels", *Australian Journal of Linguistics* 17, 155-184.

Hillenbrand, J. M. (2013). "Static and dynamic approaches to vowel perception", in Morrison, G. S. and Assmann, P. F. (eds), *Vowel Inherent Spectral Change* (Berlin-Heidelberg: Springer Verlag), 9-30.

Jacewicz, E. and Fox, R. A. (2012). "The effects of cross-generational and cross-dialectal variation on vowel identification and classification", *Journal of the Acoustical Society of America* 131, 1413-1433.

Morrison, G. S. (2013). “Vowel inherent spectral change in forensic voice comparison”, in Morrison, G. S. and Assmann, P. F. (eds), *Vowel Inherent Spectral Change* (Berlin-Heidelberg: Springer Verlag), 263-282.

Voice quality in Panjabi English: sociophonetic and forensic considerations

Jessica Wormald, Peter French, Philip Harrison, Richard Rhodes, Christin Kirchübel, and Dominic Watt

University of York & JP French Associates

This paper explores whether speakers of a contact variety of British English in two UK cities share similar vocal settings. Vocal Profile Analyses (VPAs; Laver 1991) were completed using reading passage and paired conversation recordings of 64 speakers. As well as providing a better understanding of segmental variation and thus varieties as a whole (e.g. Stuart-Smith 1999; Trudgill 1974), voice quality has been reported as one of the most powerful tools for characterising speakers in forensic contexts (Gold & French 2011, Nolan 2005).

'Panjabi' English (PE) refers to the variety spoken by British-born individuals with at least one parent who is a native Panjabi speaker and a first-generation migrant from the Panjab region. 'Anglo' English (AE), by contrast, refers to the variety spoken by individuals with no heritage language other than English, whose parents and grandparents were born in the UK. The characteristic voice quality profiles of varieties are considered using recordings of men and women from Bradford and Leicester aged between 19 and 53.

Modified VPAs were completed for each speaker in two conditions: reading passage and paired conversation. The VPA (modified for forensic use by J P French Associates) excludes the prosodic and temporal elements included in the original protocol and uses a three, rather than six, point scale (Slight, Marked, and Extreme). The VPA analysis is primarily an auditory exercise, with acoustic inspection used to complement and further explore auditory impressions. VPAs were completed by the first author, with other authors completing subsets independently as a checking mechanism.

As well as presenting descriptive results that offer valuable reference data for forensic casework, the paper explores whether variation is greater between different language backgrounds or regional groups, and whether individual variation is stronger than any group differences. The prevalence of individual settings and combinations within the dataset is illustrated in respect of their potential forensic value. Ultimately, we emphasise the need to go beyond the segmental level when describing language varieties and assessing the similarity of different voice samples.

References

Gold, E. & French, J.P. (2011). International practices in forensic speaker comparison. The International Journal of Speech, Language and the Law, 18(2), 293-307.

Laver, J. (1991). *The Gift of Speech: Papers in the analysis of speech and voice*. Edinburgh: Edinburgh University Press.

Nolan, F. (2005). Forensic Speaker Identification and the Phonetic Description of Voice Quality. In: Hardcastle, W. J. and Mackenzie Beck, J. (eds). *A Figure of Speech: A Festschrift for John Laver*. London: Lawrence Erlbaum Associates, pp. 385-411.

Stuart-Smith, J. (1999). Glasgow: Accent and Voice Quality. In: Foulkes, P. and Docherty, G. eds. *Urban Voices*. London: Arnold, pp. 201-222.

Trudgill, P. (1974). *The social differentiation of English in Norwich*. Cambridge studies in linguistics,. Cambridge: Cambridge University Press.

Can phonetically denasalised nasals trigger Korean regressive nasal assimilation?

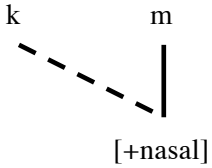
Kayeon Yoo (University of Cambridge)

Korean has a well-established phonological rule that causes coda obstruents to become nasalised before a nasal onset. Jun (1996) claims that the domain of application of this regressive nasal assimilation rule is the Intonational Phrase. This rule potentially interacts with the recently established phenomenon of domain-initial denasalisation in Korean (Kim, 2011), a process whereby onset nasals are realised with no nasality and, sometimes, with no voicing, a burst and/or aspiration (Yoo, 2015). The process targets nasal consonants that occupy the initial position of a (Prosodic) Word or higher prosodic domain (Yoshida, 2008) or an Accentual Phrase or higher domain (Kim, 2011; Yoo, 2015). By dint of the two processes just described, nasals in AP initial position (and Prosodic Word initial position, according to Yoshida) are expected to be involved in both denasalisation and regressive nasal assimilation. IP-initial nasals will not trigger regressive nasal assimilation, as this does not occur across IP boundaries.

This paper investigates how denasalisation and regressive nasal assimilation in Korean interact, by conducting auditory and acoustic analyses of recordings of 14 speakers of Kyungsang Korean containing the sequences /hanku**k**#**m**al/ 'talk (about) Korea' and /hanku**k**-**m**al/ 'Korean language'. The latter is used as a control in which the word-internal /m/ does not undergo denasalisation and triggers regressive nasal assimilation to yield [haŋguŋmal]. On the other hand, the surface form of /hanku**k**#**m**al/ will be examined to determine which of the following is found:

(a) [haŋguŋbal], by a counterbleeding ordering of nasal assimilation before denasalisation
(b) [haŋgukbal], by (i) a bleeding order of denasalisation before (non-application of) nasal assimilation, or (ii) a feeding order of nasal assimilation followed by denasalisation to the linked structure formed as a result of nasal assimilation (see Figure 1.), and
(c) [haŋguŋmal], by (i) a bleeding order of nasal assimilation followed by a failure to apply denasalisation to only a part of a linked structure formed, or (ii) an inconsistent application of denasalisation by some speakers, given the finding that Kyungsang Korean has a weaker tendency for densalisation than Seoul Korean (Yoshida, 2008).

Figure 1. Formation of a linked structure by regressive nasal assimilation



The results will have theoretical and descriptive implications. If the form in (a) is attested, this would provide strong evidence for the existence of the feature [+nasal] at an abstract level of representation. In addition, the relative ordering of domain-initial denasalisation can be determined in relation to regressive nasal assimilation and contribute to the understanding of the nature and status of domain-initial denasalisation in Korean.

Jun, S. A. 1996. Ch4.2. Experiment 2: the domain of obstruent nasaslisation. In *The phonetics and phonology of Korean prosody: intonational phonology and prosodic structure* (pp.103-125). New York, NY: Garland.

Kim, Y. S. 2011. *An acoustic, aerodynamic and perceptual investigation of word-initial denasalization in Korean*. Doctoral dissertation. University College London.

Yoo, K. 2015. *Domain-initial denasalisation in Busan Korean: a cross-generational case study*. Proceedings 18th ICPhS Glasgow.

Yoshida, K. 2008. *Phonetic implementation of Korean "denasalization" and its variation related to prosody*. IULC Working Papers Online. Vol. 8. Bloomington: Department of Linguistics, Indiana University.