CITY UNIVERSITY
UNIVERSITY COLLEGE LONDON
UNIVERSITY OF WESTMINSTER

BAAP2010
COLLOQUIUM
LONDON

British Association of Academic Phoneticians

Abstracts and Colloquium Handbook

London, March 29-31, 2010

The Conference Venue

The history of 309 Regent Street, the listed headquarters building of the University of Westminster, dates back to the1830s when it was built as part of the Royal Polytechnic Institution.
In the 1840s, the first photographic studio opened on the roof and in the 1870s it became famous for its magic lantern shows. During the 1880s, the RPI was bought by Quintin Hogg and re-opened as the Young Men's Christian Institute which in turn became known as the Regent Street Polytechnic in 1891.RSP developed and grew, changing its name again in 1970 to the Polytechnic of Central London
Twenty-two years later, in 1992, PCL in turn became the **University of Westminster**.

Within this building, the **BAAP 2010 Colloquium** will have daily use of the historic Fyvie Hall, Boardroom, Deep End and Old Cinema.

Registration on the first morning will take place in the foyer, accessed via two of the oldest revolving doors in London. This large marble stone space remains virtually unchanged from its original design.

Oral presentations will take place in the oak-panelled and stained glass
Fyvie Hall.
Here too, refreshments will be available on arrival and tea and coffee will be served during the day.

Your sandwich lunch will be served in
The Deep End.
Adjacent to Fyvie Hall, recently restored and refurbished, and still retaining a number of original features, the Deep End is the location of London's first swimming pool. This space will be available for your use throughout the day.

The BAAP 2010 Plenary Lecture, and the film show will both take place in
The Old Cinema
(the location of the first 'cinema' in London where, in February 1896 the Lumière brothers screened the first UK showing of their earliest motion film).

Finally, posters and the wine reception will take place in the
Boardroom
reached via the main staircase, on the first mezzanine floor.

FACILITIES

Toilets

Toilets are located in the basement. Further toilet facilities can be found on the upper floors, opening off the main staircase, ladies and gentlemen on alternate floors. Please ask for directions.

There is no cloakroom facility and delegates are advised to keep their personal belongings with them.

Fire Safety

In the event of a fire alarm (a continuously ringing bell) please leave quickly, via the nearest signposted exit to the designated assembly point (the corner of Cavendish Square and Mortimer Street, at the back of the building).

Leaving from the front of the building turn immediately left and take the first left at the traffic-lights outside EAT. The assembly point is at the next corner.

ABSTRACTS

Monday 29th March 2010

L1 attrition of the lateral phoneme /l/

Esther de Leeuw[1], Ineke Mennen[2] & James Scobbie[3]
University of the West of England[1], Bristol, Bangor University[2], Queen Margaret University, Edinburgh[3]

Acquisition of a second language (L2) in a migrant setting, combined with limited contact to a person's native language (L1), may lead to a decrease in ability to access one's native language system (de Bot, 2007; Cook, 2003; Köpke, 2007; Schmid, 2002). This phenomenon is referred to as L1 attrition (henceforth attrition). We examined a phonetic dimension of attrition by investigating the lateral phoneme /l/ in ten German native speakers living in Anglophone Canada. The late consecutive bilinguals had acquired German in a monolingual environment, and moved to Canada between the ages of 16 and 32, when L2 acquisition began. At the time of recording, they had been living in Canada for between 18 and 55 years.

It is generally accepted that the lateral phoneme /l/ is "dark" in Canadian (and American) English (Ladefoged & Maddieson, 1996; Olive, Greenwood, & Coleman, 1993; Wells, 1982), whereas in Standard German it is "light" (Kufner, 1970; Moulton, 1970; Wells, 1982). Recasens (2004, p. 594) states that, "dark [l] has been set in contrast with clear [l] based on well defined articulatory and acoustic properties, namely, the formation of a post-dorsal velar or pharyngeal constriction and active pre-dorsum lowering causing F2 to lower and F1 to raise". In the present research, an initial analysis of two control groups substantiated that the frequency of F1 in the lateral phoneme /l/ of the German control group was significantly lower, and the frequency of F2 significantly higher, than that of the Canadian English control group (see Table 1).

		German Control	Bilinguals in German	Bilinguals in English	English Control
Females (n=7)	F1 (Hz)	348.5 (42.0)	428.6 (84.2)	505.8 (113.0)	548.7 (81.8)
	F2 (Hz)	1863.6 (197.5)	1823.9 (228.2)	1396.3 (275.6)	1061.14 (146.1)
Males (n=3)	F1 (Hz)	244.2 (31.8)	389.9 (43.2)	442.9 (50.9)	469.5 (74.7)
	F2 (Hz)	1551.2 (72.1)	1343.7 (420.7)	988.2 (184.8)	891.4 (82.7)

Table 1: F1 and F2 (Hz) in the lateral phoneme /l/ preceded by the high front vowel /i/. Mean and standard deviations from all tokens in each group.

The investigation of the late consecutive bilingual migrants revealed that attrition in the lateral phoneme /l/ was more clearly evidenced in the frequency of F1 than in the frequency of F2. In the group analyses, all tests indicated attrition in the frequency of F1 for female and male participants, or a significant difference between the German control group and the German of the bilinguals. In contrast, the frequency of F2 did not significantly differ between the female migrants and their respective control group, although the male F2 frequency did. In the analyses of individual participants, only two late consecutive bilinguals appeared to evidence attrition in the F2 frequency of their German /l/, whereas eight (the same two plus six more) did so in the frequency of F1. Two female participants evidenced no attrition in the lateral /l/. In sum, the findings suggest that place of constriction was less prone to attrition than openness, or pre-dorsum lowering. If the former is associated with F2 frequency, and the latter with F1 frequency, it appeared that lack of post-dorsal velar or pharyngeal constriction was maintained more often than a relatively high pre-dorsum position in the German /l/ of the late consecutive bilinguals.

Voicing, devoicing and glottalisation - patterns of stop production among German speakers

Elke Philburn
University of Manchester

Investigations into German connected speech processes have shown that stop glottalisation can occur in place of a nasally released stop when the stop is followed by a nasal and preceded by a nasal or, less frequently, by a vowel or /l/ (Kohler 1994,1995,1996). Glottalisation can also occur before /l/ but is less common. This paper argues that the presence or absence of glottalisation in stop-nasal contexts may be associated with tendencies of passive stop voicing or devoicing in other voiced environments. Laryngograph recordings of eight speakers of Standard North German were taken, containing the fortis stops /p,t,k/ and the lenis stops /b,d,g/, each preceded by vowels, nasals or /l/, and followed by nasals or /l/. The data were analysed using a custom-made tool for measuring vocal fold contact duration in the Laryngograph waveform.

In the stop-nasal contexts, two ‘groups’ of subjects emerged: Four of the eight subjects produced glottalized variants to varying extent, while the other four produced nasally released stops throughout. In the stop-lateral contexts, no glottalisation occurred with any of the eight subjects, but a pattern did emerge that appeared to confirm the division of subjects into two groups. The four subjects who glottalised stops in stop-nasal contexts showed a tendency towards passive voicing of the fortis stops, while no such tendency was found with the remaining four speakers, who produced fully devoiced stops throughout. For the lenis stops, three of the four subjects who had not glottalised stops in stop-nasal contexts showed a tendency towards passive devoicing, while no such cases occurred with the other group.

Although drawing on a limited number of speakers, the findings suggest a possible connection between the employment of stop glottalisation and the occurrence of passive fortis stop voicing. Several instrumental investigations have shown that nasally released stops are associated with lower velic height than orally released stops (Kohler 1977,1980; Künzel 1979). Kohler (1977) found that for the production of nasally released stops, activities of the palatoglossus and the levator palatini muscles can interfere with each other, which may inhibit velic elevation and in some cases may result in the elimination of a stop articulation. The employment of glottalisation, on the other hand, allows for a reduction or even omission of velic displacement and is assumed to increase articulatory ease (Kohler 1994,1995,1996).

With regard to the present findings, it is possible that a low velum position might facilitate passive voicing, as an insufficient closure of the velopharyngeal port may inhibit the stoppage of airflow during the oral closure. As four of the eight subjects did not employ glottalisation but produced nasally released stops, it may be the case that these subjects were less affected by the observed limitations of velic height control and hence achieved sufficient velic height throughout. If this is a characteristic that applies to stop production in more general terms and not only to nasally released stops, an effect on stop voicing or devoicing in other voiced environments would be imaginable. If, for example, speakers displayed a tendency towards increased velic height in stop production, this could facilitate the production of nasal releases while at the same time facilitating a stoppage of airflow and a decrease in transglottal air pressure which could result in passive devoicing. The paper calls for an analysis of a more extensive body of data and additional instrumental techniques to shed more light on this observed phenomenon.

What can epiphenomenal sound production tell us about the production of ‘true’ ejectives?

Adrian P. Simpson
Universität Jena

The temporal overlap of the phonetic correlates of adjacent phonological elements can give rise to sounds which are acoustically and, in many cases, audibly discernible from the phonetic correlates generally associated with the phonological elements they have emerged from.
Non-pulmonic epiphenomenal sound production is of particular interest, not least in relation to processes of sound change. Ohala (e.g. 1995, 1997) has described and exemplified possible mechanisms behind some types of epiphenomenal clicks and ejectives, and studies since Marchal (1987) have shown that such emergent sound production is a widespread feature of a number of languages not generally associated with non-pulmonic sound production in their phonologies.

While Ohala’s work has uncovered possible mechanisms behind many types of epiphenomenal sounds, some details of the patterns we find cannot be adequately accounted for in the terms he proposes. Epiphenomenal ejectives in German are an example of this (Simpson 2007). The overlap of a final plosive and junctural glottalisation in a vowel-initial syllable in German, e.g. [veːtʔaɪn] *weht ein,* can give rise to a plosive release having the auditory and acoustic characteristics of an ejective. One account which has been offered for this (Ohala 1997, Simpson 2007) is articulatory movement occurring once the double oral and glottal closure has been made. So, for instance, any vowel-vowel movements flanking the plosive will change the volume of the supraglottal cavity with a subsequent increase or decrease in intraoral pressure. The subsequent release of the plosive will therefore be fuelled by an ingressive/egressive glottalic air stream. However, the strength of many glottalically fuelled plosive releases suggests that such volume changes might not be sufficient to give rise to the required change in pressure. An alternative account is that strong glottalically fuelled releases are not epiphenomenal, but are rather the result of an active upward movement of the larynx. However, there is another epiphenomenal account. Although the glottis is closed or predisposed for creak on release of the plosive, the necessary build up of pressure during the closure phase of the plosive can be accomplished with a pulmonic air stream. The plosive release is then strictly speaking fuelled by pulmonic air, but furnished with the auditory and acoustic quality of an ejective and, most importantly, no change in supraglottal cavity volume is required.

While providing a possible explanation of epiphenomenal ejective bursts in a language like German, this account raises the possibility of a similar mechanism lying behind at least some of the phonological ejectives in the world’s languages. The role played by larynx movement in providing adequate pressure change has been drawn into question (e.g. Kingston 1985 for Tigrinya). Furthermore, a pulmonic component has been attributed to ‘voiced ejectives’, seen to rather be a sequence of voiced pulmonic plosive followed by a regular ejective, i.e. [dt’] (Snyman 1970, 1975).

Using acoustic data from both epiphenomenal ejectives in German and phonological ejectives in Georgian together with acoustic and EGG data of final ejectives in Suffolk English the plausibility of pulmonically fuelled ejective releases is discussed.

Syllable frequency effects on coarticulation and short-term motor learning

Frank Herrmann, Stuart Cunningham & Sandra Whiteside
University of Sheffield, Department of Human Communication Sciences

Psycholinguistic research suggests that articulatory routines for High frequency syllables are stored in form of gestural scores in a syllabary (e.g. Levelt & Wheeldon, 1994). Syllable frequency effects on naming latency and utterance duration have been interpreted as supporting evidence for such a syllabary (Cholin, Levelt, & Schiller, 2006). Whiteside & Varley (1998a, 1998b) hypothesized that high frequency verbo-motor patterns and articulatory sequences would also result in greater consistency and greater degrees of articulatory overlap (coarticulation) in the speech signal.

This paper presents the acoustic analysis of a data-subset from a project investigating speech motor learning as a function of syllable type. Twenty-four native speakers of English were asked to listen to and repeat 16 monosyllabic stimuli which belonged to either of two categories (High vs. Low frequency syllables) and had word status (Herrmann, Whiteside & Cunningham, 2009). In addition, 16 disyllabic non-word stimuli were generated using the monosyllabic stimuli as components (e.g. *boost* & *dot* => *dot.boost*).

Significant syllable frequency effects were found for the durational and coarticulation measures. High frequency syllables exhibited greater degrees of coarticulation and greater overall consistency in their production than Low frequency syllables irrespective of their context (i.e. 1st or 2nd position in a disyllabic non-word or as a monosyllabic word).

An analysis of short term learning across ten repetitions revealed significant differences between the two syllable frequency categories for both durational and coarticulation measures. High frequency syllables showed greater degrees of coarticulation and greater consistency across ten repetitions than Low frequency syllables, which varied in their degree of coarticulation.

These data provide some further supporting evidence that different syllable categories may be encoded differently during speech production.

The Manchester Polish STRUT – investigating the acquisition of local accent features in the speech of non-native English speakers living in Manchester

Rob Drummond
University of Manchester

When non-native speakers of English are exposed to a variety of English which is different from that with which they are familiar (be that a pedagogical model or a notion of a 'standard' dialect), certain features of this variety can be acquired into their own speech. My research addresses this topic by investigating the extent to which the pronunciation of Polish people living in Manchester, who are using English as a second language, is influenced by the local accent.

Several features of the local accent are being investigated (t-glottaling, h-dropping, -ing), but this paper will focus on one feature in particular - the vowel sound in STRUT words. The STRUT vowel is a highly salient feature of Northern British English, and one which generally shows little, if any, contrast with the FOOT vowel in the speech of native English speakers in the Manchester area. Those native speakers who do show some contrast tend to produce a schwa-like sound for STRUT in some contexts, but rarely produce anything close to RP STRUT (Wells 1982). In contrast, the Polish speakers being investigated have all been exposed to a pedagogical model involving something similar to RP STRUT, and the Northern English variety represents a deviation from this model.

The data presented in this paper represent emerging patterns of acquisition in the speech of 30 participants aged between 18 and 40, all of whom had some level of English when they arrived in the UK. Speech data have been gathered by recording informal interviews and attitudinal data by using a questionnaire. Participants' awareness of the local accent has also been measured in a matched guise perception task. Relevant STRUT tokens have been analysed both auditorily and acoustically in order to get a fuller picture of any change.

Those participants who have been in Manchester for only a short time show complete consistency with the pedagogical model, i.e. something close to RP STRUT. Unsurprisingly, as the Length of Residence (LOR) increases, so does the likelihood of there being some degree of change in the realization of the vowel. However, this only gives part of the story, as other factors such as gender, desire to integrate and type of exposure to the local accent also have an influence on the degree of acquisition. This paper will also comment on the emerging pattern of acquisition from a lexical point of view, and offer suggestions as to why some words are produced with the local variant before others.

Accent morphing in spoken language conversion: preserving speaker identity whilst changing their accent

Kayoko Yanagisawa & Mark Huckvale
Department of Speech Hearing and Phonetic Sciences, UCL

Spoken language conversion (SLC) aims to generate utterances in the voice of a speaker but in a language unknown to them, using speech synthesis systems and speech processing techniques. It has applications in speech-to-speech translation systems, dubbing of foreign language films, and foreign language learning. Previous approaches have been based on voice conversion (VC), and used training and conversion procedures based on statistical models to change the speaker characteristics of a given speech signal from the output of a foreign language (L2) text-to-speech (TTS) system to the target speaker (e.g. [1], [2]). Application of this technique across languages has underlying assumptions which ignore phonetic and phonological differences between languages, and thus lead to a reduction in the intelligibility of the output [3].

A new approach to SLC, accent morphing (AM), was previously proposed and evaluated in terms of intelligibility [4]. Rather than taking a second speaker and modifying their voice characteristics as in VC, AM attempts to preserve the characteristics of the target speaker whilst modifying their accent, using phonetic knowledge obtained from a native L2 speaker. It takes parallel utterances, one by the target speaker speaking L2 with a foreign (L1) accent, and the other by an L2 speaker which may be a TTS system. The target speaker's L2 utterance with L1 accent could be generated using an L1 TTS. It then uses audio morphing to change those aspects of the target speaker speech signal which are related to accent, to modify the accent characteristics towards the native L2 speaker's, whilst preserving as much of target speaker characterisitics as possible, such as high frequency spectral envelope, voiceless frames, speaking rate and mean F0. It was shown in the above study that AM is capable of improving the intelligibility of foreign-accented speech.

In the present study, we investigated the performance of AM and VC systems in terms of the extent to which the output sounded like the target speaker. Various AM conditions were tested, to see if it was possible to preserve more of target speaker characteristics by morphing selectively in the time domain and in the frequency domain. In a listening test with 45 L2 listeners, the best AM condition achieved a target speaker similarity rating of 4.39 (1=completely different, 7=identical) whilst VC was given a rating of 4.03. The two systems were comparable in terms of intelligibility. Examination of the various AM conditions revealed that it was possible to generate output sounding more like the target speaker by morphing selectively. This is an advantage over VC, which requires signal processing to be applied to the entire utterance to be converted, thereby introducing artefacts. It was found, however, that selectively morphing reduced the intelligibility of the output, exposing a trade-off relationship between speaker identity and intelligibility in SLC.

Is the NURSE vowel in South Wales English a front rounded vowel? An acoustic and articulatory investigation

Robert Mayr
Centre for Speech and Language Therapy, University of Wales Institute Cardiff

Impressionistic accounts of South Wales English (SWE) suggest front rounded realizations of the vowel in the NURSE lexical set (e.g., Collins & Mees, 1990; Mees & Collins, 1999; Penhallurick, 2008; Walters 1999, 2001; Wells, 1982). However, the specific phonetic properties of the vowel are not described uniformly in these studies. Moreover, they have relied entirely on auditory descriptions, but do not involve instrumental analyses. The study presented here is the first to provide a systematic acoustic account of the spectral and temporal properties of the NURSE vowel in South Wales English, coupled with an articulatory investigation of its lip posture from frontal and lateral views. The study also explores the relationship of the vowel to realizations of the same phoneme in Standard Southern British English (SSBE), and to those of the long close-mid front rounded vowel of Standard German (SG).

The results indicate systematic differences between the three vowels, with the SWE vowel produced with an open rounded lip posture, but the acoustic properties of an unrounded front vowel. This could suggest that the SWE vowel is indeed a front rounded vowel, as auditory descriptions have claimed. After all, the acoustic analysis suggests a front vowel and the articulatory analysis a rounded vowel. However, in this paper it will be argued that for a vowel to be considered ‘front rounded’, it is not only required to be ‘front’ and ‘rounded’, but its articulatory gesture also needs to result in the characteristic lowering of F2 and F3 frequencies. This is the case for the SG vowel and front rounded vowels in languages across the world (e.g. Gendrot & Adda-Decker, 2005; Linker, 1982; Pols, Tromp & Plomp, 1973; Wood, 1986), but crucially not for the NURSE vowel in SWE. Perhaps there is a tipping point in the lip gesture dimension which the SWE vowel has not reached. On the basis of the articulatory and acoustic data it is argued that the SWE vowel is best represented as a slightly rounded long close-mid front monophthong, with [e] used as a base symbol.

A Phonetic Study of Ramsau am Dachstein German

Elfriede Tillian & Patricia Ashby
Dept of English, Linguistics and Cultural Studies, University of Westminster,

This poster presents a descriptive phonetic study of the South/Middle Bavarian Austrian accent of German, spoken in the area of Ramsau am Dachstein (RADG), Styria. Based on the speech of an 88 year-old female subject (who also contributed the translation of *The North Wind and the Sun* into her dialect), this study offers tabulation and description of consonantal and vowel features alongside description of apparent allophonic processes, suprasegmental features and processes of connected speech visible in the data. Findings are compared where relevant to the phonetics of Standard German (SG) [KOHLER 1977, 1989]. For some 75% of Austrians, non-standard dialect is the language of daily life [WIESINGER 1990]. The diglossic situation typical of neighbouring Switzerland does not pertain. The speech examined here is typical of a disappearing variety [PERNER 1972], untouched by the influences that impact on the speech of younger generations.

Features of particular interest in this variety of German include a strong preference for voiceless articulations in the obstruent series which in turn gives rise to an apparent absence of voicing contrasts. For example, compared to SG, RADG has no voicing contrast in utterance initial plosives (no voicing is detected at all, and no aspiration at least for bilabials and alveolars). RADG also includes velar affricates, uvular fricatives (but no palatal ones – so, no typical SG *ich*-laut/*ach*-laut patterning), and an unstable rhotic (with manifestations including, among others, spirantized and vocalized variants – the latter not unlike SG post-vocalic /r/).

In addition, vocalization processes (including possible intervocalic l-vocalization) give rise to large numbers of diphthongs. Among the monophthongal vowels, there is an absence of clear-cut length definition (again in contrast to SG). Front rounded vowels are conspicuously absent in RADG. Comparison with Standard German therefore reveals striking differences in both the vowel and consonant systems.

The poster also considers very briefly the apparent absence of the characteristic 'sing song' rise-fall intonation that is a recognized hallmark of Styrian speech [WIESINGER 1967].

Acoustic quantification of aspiration: a standardization based on analysis of preaspiration in full phonation and whisper

Olga Gordeeva [1] & James M. Scobbie [2]
Acapela Group, Mons, Belgium[1]; Queen Margaret University, Edinburgh [2]

Acoustic quantification of aspiration before word-final obstruents (other than duration) has received little attention in the phonetic literature (see e.g. Ní Chasaide and Gobl, 1993; Bombien, 2006). One reason for this is that spectral estimates of phonatory voice settings (e.g. modal vs. whispered) which rely on formant levels (e.g. spectral tilt in Hanson, 1997) do not differentiate between contributions from periodic or aperiodic excitation (Ní Chasaide and Gobl, 1993: 320). Another reason is the occurrence of variation in high spectral energy due to large differences in supra-laryngeal friction accompanying preaspiration, ranging from more anterior to more posterior pharyngeal or laryngeal places of articulation (e.g. [ç x h]) (Laver, 1994; Silverman, 2003). Finally, most acoustic measures of aspiration used to date are periodicity-dependent: i.e. they are not computable in non-periodic portions of aspirated transitions (e.g. Hillenbrand, Cleveland, and Erickson, 1994). The latter is very problematic, because such portions are commonly occurring in preaspirated obstruents.

In a number of recent studies (Gordeeva and Scobbie, 2007, to appear) exploring the linguistic functioning of preaspiration in Scottish English word-final fricatives, we developed and used a new *periodicity-independent* acoustic measure of aspiration derived from the standard zero-crossing rate in the time-domain. The performance of this measure was compared to a set of more established (but periodicity-dependent) acoustic correlates of aspirated phonation: i.e. the ratio between the spectral levels of the first and second harmonics (H1-H2, Hanson, 1997), and harmonics-to-noise ratio. The conclusion from our studies was that a single periodicity-independent measure is better able to quantify linguistic use of aspiration in either fully-phonated or whispered voice (or any states in-between), than the periodicity-dependent ones.

The aim of this study is to strengthen the methodological base for the use of our periodicity-independent quantification of aspiration. We: (1) compare preaspiration in normally phonated productions vs. fully whispered productions (N tokens = 2251) in a group of five linguistically naïve working class speakers from Edinburgh; (2) relate the highest measured rates of zero-crossings (i.e. high levels of aspiration) to the auditory annotation (perception) of preaspiration in the normally phonated tokens. We use data from Scottish English because the phonological distinction between word-final “voiced” and “voiceless” fricatives is conveyed in part by the duration of phonation vs. strong aspiration/whisper in vocalic parts before the fricative. We report the ranges of preaspiration in terms of zero-crossings valid against whispered phonation and the speaker-independent ranges. We also discuss the implications for this method for phonetic studies.

An acoustic and auditory analysis of glottals in Southern British English

Joanna Przedlacka & Michael Ashby
University College London

This paper revisits data from teenage (14-16 years old) speakers of British English, covering the Home Counties, RP and East London, which had previously been gathered for a sociophonetic study employing an auditory analysis. The newly augmented data consists of nearly a thousand tokens of syllable non-initial /t/, from 22 speakers, where glottal or preglottalised variants are possible realisations. The recordings, having been made in the field, are of variable quality, but as far as possible have been subjected to acoustic analysis. A first finding, in line with previous reports concerning other varieties of English (Hillenbrand and Houde 1996, Docherty and Foulkes 1999), is that the 'canonical glottal stop' consisting of an interval of silence produced by a closed glottis is essentially non-existent. The majority of tokens in the present study had no clear gap, but a range of realisations such as continued vowel formants, differences in voicing, creak or periods of irregularity during the neighbouring vowels. It is found that various degree-of-voicing measures (Holmes, 1998), especially the autocorrelation function, show well-defined local minima in the region where the glottal constriction is expected, permitting the annotation of files for automatic analysis. A range of parameters, including fundamental frequency, intensity, autocorrelation function and zero-crossing rate can be extracted to give an acoustic profile of the glottal events. The observation that the degree-of-voicing measurements are at a minimum in this region goes some way towards answering the question why despite vocal fold vibration the perception of the glottal stop is voiceless.

Clicks in York English

Richard Ogden
Department of Language & Linguistic Science, CASLC, University of York

Clicks are frequent in spoken English, but do not form parts of words. A typical textbook description of clicks in English, based on hunches about meaning is: ‘the interjection expressing disapproval that novelists write “tut-tut” or “tsk-ts”,'' (Ladefoged 2001: 119). Although there are numerous claims in the literature that clicks in English express things like disapproval, irritation or annoyance (e.g. Clark & Yallop 1990, Gimson 1970, Ladefoged 2001, Laver 1994, Ward 2006), more recent work by Wright (2007: 1069) shows that clicks have other functions in conversation. One such is "to demarcate the onset of new and disjunctive sequences [of talk]".

There remains very little empirical work on clicks in English. Wright's ground-breaking work is based on recordings of poor quality, making reliable acoustic analysis difficult; it has little quantification and the analysis is restricted to a couple of sequential environments. The clicks identified by Wright alternate freely with other kinds of articulation, such as the articulators coming apart and coincidentally making percussive noises on an in-breath rather than being a deliberately articulated sound.

This poster presents results from a preliminary study of York English which aims to collect more data on the frequency, nature and function of clicks in spontaneous speech. The data were taken from an ESRC project, “A Comparative Study of Language Change in Northern Englishes”, a sociolinguistic study of varieties of Northern English. The talk of four male and four female pairs of speakers from approximately 280 minutes of spontaneous speech in sociolinguistic interviews was studied. Clicks, ejectives and percussives were all identified.

In all, 222 definite clicks were identified in the data, i.e. in every five minutes of talk, approximately four clicks occurred. However, a number of difficulties were encountered with the phonetic analysis, making the interpretation of these data not straightforward:

- velarically initiated suction stops are not always easily distinguished from pulmonically initiated ingressive percussives
- place of articulation is not always easily identified
- loudness, plosion type and duration are not always easy to identify

The interactional functions and sequential locations of the data were also considered. While the findings are not conclusive, they suggest:

- clicks are implicated in many more sequential environments and in indexing social actions than has previously been thought
- in some places in sequence (e.g. in a display of sympathy), ‘clicks’ must be realised as velarically initiated suction stops; in other places (e.g. turn-initially), ‘clicks’ can be articulated in a wide range of ways

The distribution of clicks appears highly personal: some speakers seem to be frequent ‘clickers’ and others do not. There is an apparent effect of gender, with far more clicks produced by females than males. However, this is not yet a robust claim because transcribers’ judgements are not always in agreement; and the function of clicks is not clear in all cases, so it may be that people who appear to click frequently are speakers who do the kinds of actions that make the production of a click relevant.

The poster presents some of the findings from the study of York English, and raise more general questions about clicks: are they sociolinguistic markers? are at least some of them the results of articulatory ‘gearing up’? how can the different types of ‘click’ best be analysed in phonetic terms? Suggestions for further work are made.

Modification of phrase-final tone targets under time pressure: Evidence from the intonation systems of German and Russian

Tamara Rathcke[1] & Jonathan Harrington[2]

Department of English Language, University of Glasgow[1], Institute of Phonetics and Speech processing, Ludwig-Maximilians-University of Munich[2]

Two strategies have been proposed to account for the effect of time pressure on realisation of phonological tones: the intended tonal pattern can be either (1) produced completely in a shorter period of time and therefore *compressed* as in Fig. 1.A, or (2) realised incompletely causing a target undershoot called *truncation* as in Fig. 1.B (Erikson & Alstermark, 1972; Bannert & Bredvad, 1975; Grønnum, 1989; Grabe, 1998). Two phonetic features are incorporated in this model: target approximation (TARGET) and rate adjustment (RATE). So, compression can be also described as [+TARGET; +RATE] as opposed to truncation which is [-TARGET; -RATE]. It has however been recognised in previous studies that these two strategies are insufficient for modelling the variability found in the production data (Grabe et al., 2000; Hanssen et al., 2007). Obviously, the model is unsystematic in omitting the possibility of at least two additional strategies, i.e. [-TARGET; +RATE] and [+TARGET; -RATE]. There is some evidence that rate adjustment without a proper target approximation (i.e. [-TARGET; +RATE] as *accomodation* in Fig 1.C) is used in some accents of British English (called „range compression" in Grabe et al., 2000) as well as in Dutch (Hanssen et al., 2007). A good target approximation without any rate adjustment, i.e. [+TARGET; -RATE], implies the kind of temporal reorganisation of the tonal pattern as indicated by *compensation* in Fig 1.D. Numerous studies in the AM framework of intonation have shown that this strategy is very common in intonation languages (e.g. Prieto, van Santen & Hirschberg, 1995; Schepman, Lickely & Ladd, 2006; among many others).

Based on this model, we carried out a study on the realisation of H+L* and L*+H pitch accents with a low boundary tone in German and Russian. We devised sets of materials in which the phonological structure of phrase-final accented words was varied systematically by shortening the amount of material available for voicing from relatively long words (German *Linner*; Russian *Kalinkin*) to extremely short words with nuclei flanked by voiceless consonants (*Schiff*; *Rashif*). In general, the results confirmed the need for the four-way distinction in Fig. 1. For example, German falls corresponding to H+L* pitch accents were distinguished by accommodation (Fig. 1C) in which the low target of L* was preserved as compared to truncation (Fig. 1.B). We also found that rising L*+H pitch accents were reorganized temporally under time pressure in accordance with compensation (Fig. 1.D) in both German and Russian.

The general conclusions from these results are firstly that compression and truncation alone are insufficient to account for tonal modification under time pressure, secondly that such phonetic adjustments are crucially sensitive to the composition of the phonological tonal string, and thirdly that the strategies for adjusting phonological tones under time pressure are language specific.

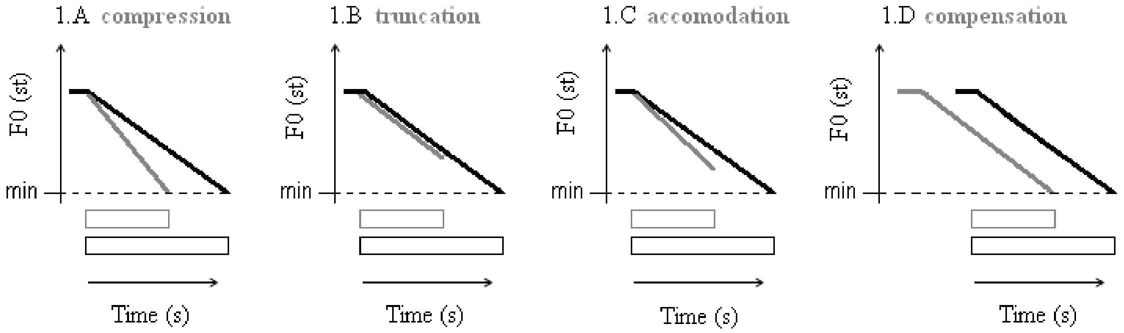


Fig.1. *Four types of tonal modification in two environments: without time pressure (black lines) vs. with time pressure (grey lines).*

Voice Onset Time in Serbian: distribution and variability

Mirjana Sokolović
University of Newcastle upon Tyne

There is considerable research on the factors that have been found to introduce variability in the phonetic realisation of the voicing contrast, with Voice Onset Time (VOT) receiving more attention than any other correlate of voicing. A number of factors have been found to influence VOT, such as place of stop articulation (Lisker & Abramson 1964, Docherty 1992, Jessen 1998, Cho and Ladefoged 1999, Abdelli-Beruh 2009), the quality of the following vowel (Klatt 1975, Smith 1978, Docherty 1992, Morris et al. 2008), speaking rate (Kessinger & Blumstein 1997, 1998; Miller et al. 1986, Volaitis & Miller 1992), utterance type (Lisker & Abramson 1964, 1967), age (Sweeting & Baken 1982, Ryalls et al. 1997, Ryalls et al. 2004), gender (Ryalls et al. 1997, Whiteside & Marshall 2001, Whiteside et al. 2004, Robb et al. 2005), and speaker identity (Allen et al. 2003, Theodore et al. 2009). However, we are still short of understanding this variability cross-linguistically. Most of the previous research has concentrated on languages with two-category contrast between short-lag and long-lag VOT, especially English, while languages which contrast voicing lead and short-lag stops have received less attention. The aim of the present study is to address this by investigating a language from the latter group - Serbian.

This poster presents VOT data from an acoustic analysis of the speech of twelve native speakers of Serbian. The speech sample consisted of CVC(C) words with initial stops, spoken in isolation and in a sentence frame. Since this is the first substantial study on VOT in Serbian, it first reports on the VOT distributions for Serbian stops. It then provides a quantitative account of the following linguistic and demographic factors that are known to introduce variability in VOT in other languages: stop place of articulation, quality of the following vowel, utterance type, age, gender and speaker identity. It further examines language-specific details of this variability as well as their relevance in the context of research on other languages.

On lenition in Brazilian Portuguese

Thaïs Cristófaro Silva[1,2,3] & Maria Cantoni [1,4]
UFMG[1], CNPq[2], FAPEMIG[3], CAPES[4])

Lenition involves the weakening of segments. Usually when undergoing lenition a given segment loses some property and is realized as a weaker or less strong one. In the extreme cases the segment which undergoes lenition is deleted. This paper intends to consider some synchronic cases of lenition in Brazilian Portuguese. It aims to address the gradual nature of lenition and the role played by type and token frequencies in the implementation of lenition. The first case study to be presented deals with lenition of [ks] sequences that are manifested as [s]. For example, *sinta[ks]e > sinta[s]e* “Syntax” or *tó[ks]ico > tó[s]ico* “toxic” (Cantoni, 2009). The second case study to be addressed involves lenition of affricates in postonic syllables that are lenited and realized as a stop. For example, *par[tʃ]es > par[t]es* “parts” or *tar[dʒ]es > tar[d]es* “afternoon” (Leite, 2006). Acoustic analysis indicates that the implementation of lenition is phonetically gradual (Pierrehumbert, 2001, 2003). It will be argued that lenition involves specific and targeted changes in the articulatory routines, so that new patterns are accommodated into the resulting sound (Goldstein, Byrd & Saltzman, 2006). The emergence of new or lenited pattern depends on the well-formedness conditions of the language and also on the potential accommodation of the new pattern into the lexicon of the language. Regarding frequency effects it will explored the hypothesis that phonetically based sound changes affect more frequently used words before less frequently used ones. On the other hand, sound changes that are not phonetically motivated affect less used words first (Phillips, 2001; Bybee, 2001). It will be shown that frequency effects – both type and token - are relevant in the implementation of lenition, but the lexicon has an important role to play in the organization of lenited forms. The results to be presented indicate that sound changes have a close relationship to the lexical organization of the language and also that usage is related to the emergence of new patterns.

Aspiration in Scottish Gaelic stop consonants

Claire Nance[1] & Jane Stuart-Smith[2]
Department of Celtic and Gaelic[1], Department of English Language[2], University of Glasgow

The Scottish Gaelic stop system is represented orthographically as <p t c> and <b d g>. Phonemically, these stops are aspirated /p^h t^h k^h / and voiceless unaspirated /p t k/ respectively (Ladefoged et al. 1998). Additionally each stop can be phonemically palatalised or velarised (Gillies 1993). In word-final position and post-vocalic word-medial position, the aspirated series are realised as pre-aspirated [^{h}p ^{h}t ^{h}k] (Ternes 1973). The aspiration and pre-aspiration system is similar to Icelandic (Thráinsson 1978) and to Faroese (Helgason 2003). The phonetic realisation of Scottish Gaelic pre-aspiration is reported to vary from dialect to dialect (Ó Murchu 1985; Bosch 2008).

This poster investigates the phonetic characteristics of <p t c> and <b d g>. We compare data from six native speakers of Lewis (Outer Hebrides) Gaelic across two generations: three speakers over 40, and three speakers aged 18-25. It is impressionistically reported that Scottish Gaelic phonology is changing rapidly, as might be expected in an obsolescent language (Dorian 1981, Andersen 1982). The inclusion of two generations in the sample therefore investigates the possibility of change in the realisation of the stop series. Word list data were digitally recorded in a noise-attenuated sound studio. The list was devised to include minimal pairs for all possible contrasts in Gaelic, and also to include an example from Ladefoged et al.'s (1998) word list for each phoneme in each context in order to provide comparable data. Each word was couched in a carrier phrase and the words were placed in semantic groups in order to distract attention from minimal pairs. The sentences were repeated three times in random order within the semantic groupings. The resulting recordings therefore contain examples of word initial, word medial and word final stops at all places of articulation where possible.

All tokens were labelled on the waveform for onset/offset of consonant and vowel and aspiration/pre-aspiration. A range of acoustic measures were taken including: voice onset time; a set of duration measures (cf. Jones and Llamas 2006) total voiced vowel, modal voicing, breathy voice (if any), pre-affrication (if any), total pre-aspiration (if any), voicing offset ratio (Gordeeva and Scobbie 2007, to appear); and a measure to assess noise in the signal independent of periodicity, adapted Zero Crossing Rate (cf. Gordeeva and Scobbie 2007, to appear).

New methods of pre-aspiration data analysis (Gordeeva and Scobbie to appear) applied to a language with 'normative' (Helgason 2002) pre-aspiration will provide input into our understanding of aspiration and pre-aspiration. Our examination of the Scottish Gaelic stop system as a whole (aspiration and pre-aspiration in the <p t c> and <b d g> series) will contribute to our currently extremely limited knowledge of the language's sound system. The data from two generations will also provide evidence about the current state of the language and possible changes in progress.

The effect of speaker and token variability on speech perception in children with dyslexia

Valerie Hazan, Stuart Rosen, & Souhila Messaoud-Galusi
Speech, Hearing and Phonetic Sciences, UCL

Many studies have suggested that at least a subset of individuals with dyslexia show deficits in some aspects of speech perception. It is hypothesised that poorly specified phoneme categories may lead to difficulties in making consistent sound-to-letter correspondences which in turn may make it difficult to learn to read. In our study, we explored how children with dyslexia deal with both speaker and token variability in tasks involving the identification and discrimination of naturally-produced CV syllables presented in noise. If children with dyslexia have poorly-specified phoneme categories, it would be expected that they would be more greatly affected by increased variability in the speech signal.

59 children participated in our study: 34 with dyslexia (mean age: 147.3 months) and 25 average readers (mean age: 146.8 months). For the identification tests, stimuli included the tokens /pi, bi, ti, di, fi, vi, si, zi, mi, ni, spi, sti/ presented in multispeaker babble noise. Four conditions varied in the degree of speaker and token variability: in Condition 1, all tokens were by a single speaker and produced with falling intonation; in Condition 2, tokens were by a single speaker but with falling, rising, rise-falling and steady intonation; in Condition 3, tokens were produced by four speakers with falling intonation, and in Condition 4, tokens were produced with four speakers with varying intonation patterns. Discrimination tests (using the same four graded conditions) were also run, and included the voicing (/bi/-/pi/) and place (/bi/-/di/) contrasts only.

In the identification tests, children with dyslexia made greater errors than average readers, but only in conditions with variable intonation. The introduction of speaker variability did not affect them any more than average readers. Analyses in terms of the transmission of specific phonetic features showed that this group difference in the variable intonation conditions was primarily due to a poorer perception of the voicing feature within fricatives by dyslexic children. Only about 30% of the children with dyslexia showed this pattern of errors. In the discrimination tests, children with dyslexia performed more poorly than average readers in discriminating both the voicing and place contrasts; as well as a group effect, close to 50% of children performed below norm in these tasks. Contrary to what was found for the identification tasks, the difference in scores across groups did not vary significantly across the different speaker and intonation conditions.

These results suggest that a proportion of children with dyslexia may have difficulties in perceiving a small number of phonetic contrasts in noise, and seem generally more greatly affected by the introduction of variability in intonation than average readers. The more widespread difficulties in discrimination tests across conditions suggest that these difficulties may be task-related (e.g., may be linked to the memory demands of this task). Given that the specific identification difficulties shown by children with dyslexia were mostly limited to voicing in fricatives, and that they affected a minority of children only, a causal effect on the acquisition of reading seems unlikely.

The effects of speaker accent on the comprehension of people with aphasia

Caroline Newton, Carolyn Bruce, Jane Dunton, Cinn Teng To, Hannah Berry, & Bronwen Evans
University College London

There is now a wide body of research which demonstrates that variation in both native and non-native accented speech affects speech processing in adults, with processing deficits appearing to be greater for non-native accents (see, for example, Munro & Derwing, 1995; Adank et al., 2009). Research also shows that these listeners are able to adapt rapidly to the accent of a specific speaker, and generalise this learning across speakers (e.g. Clark & Garrett, 2004; Bradlow & Bent, 2008).

The research presented here explores the effects of accent on the comprehension of individuals with a compromised language processing system following a stroke. In three separate experiments, participants with and without aphasia carried out a selection of tasks that do not require a verbal response: a sentence-to-picture matching task, a sentence verification task and a discourse comprehension task. Results indicate that individuals with aphasia make significantly more errors with unfamiliar accented speech than with a familiar accent, and that difficulties with an unfamiliar non-native accent (versus familiar and unfamiliar native accents) are more marked for these individuals than for participants without aphasia.

The findings so far, therefore, suggest that unfamiliar accent increases processing demands, and this is a particular problem for people with aphasia. The findings also raise a number of questions that we will address in future research.

Longitudinal voice monitoring using mobile devices

Felix Schaeffler & Janet Beck
Speech Science Research Centre, Queen Margaret University, Edinburgh

Voice problems and voice disorders reach almost epidemic proportions in certain professional groups, such as the teachers (e.g. Roy et al. 2004). Although this situation has been known for quite a while now, effective preventative measures are scarce, and numbers of teachers reporting voice problems continues to be high (e.g. Schaeffler & Beck, 2010).

One pre-requisite for more effective voice care and early detection of voice problems is increased knowledge about early stages of voice problems and normal day-to-day fluctuations of voice parameters. Most research into voice and health is based on cross-sectional studies, while longitudinal studies of voices under controlled but natural conditions (e.g. at the workplace) are rare. There is thus an urgent need for more work in this area.

We argue that modern mobile devices like smartphones provide the ideal platform for the recording of longitudinal voice databases under natural "field" conditions. They usually have wireless internet capabilities, allowing the user to send locally recorded data to a central server. They have quite powerful processors and allow the running of third-party software. They have advanced user interfaces (e.g. touch screens), allowing for user-friendly 'touch-of-a-button" recording software and the collection of questionnaire data alongside acoustic data, and they are sufficiently light-weight to be carried around all day.

In cooperation with a Scottish SME we have developed a software tool for the Apple iPhone and the Apple iPod Touch that allows users to record themselves for a period of up to four minutes, to fill in a voice health questionnaire associated with the recording, and to submit one recordings to a central server when a wireless internet connection becomes available.

Our current tests focus on the iPod Touch. This device shares most functionality with the iPhone (apart from the telephony part), does not require a provider contract and is reasonably priced, with a purchase price below £200. We are currently testing the audio signal quality achievable with this device using a range of different microphones, and we are assessing the reliability of acoustic voice parameters like shimmer, jitter or HNR extracted from field recordings.

Our presentation will describe the functionality of the software tool, and will also provide technical data about the signal quality of recordings made under different field conditions, together with an evaluation of the reliability of voice parameters extracted from this type of data. We assume that a modified version of our software tool could be useful for a wide range of purposes, i.e. whenever combinations of audio and questionnaire data collection are desired in field studies.

Perception of Foreign Accent Speech: Foreign, Disordered or Both?

Jana Dankovičová[1] & Claire Hunt[2]
University College London[1], North West London Hospitals NHS Trust[2]

Foreign Accent Syndrome (FAS) is a relatively rare neurological speech disorder that manifests itself by ‘foreign’ sounding speech. It emerges most frequently after stroke, but also after traumatic brain injury, and the progressive neurodegenerative diseases - multiple sclerosis and dementia.

One of the major puzzles that research has focused on to date in relation to FAS is what makes it different from other speech disorders. As yet we do not have a satisfactory answer. The perceptual impressions of ‘accent’ have been reported in most papers anecdotally rather than explored experimentally (the notable exceptions being Miller et al. 2006 and Di Dio et al. 2006), and perceptual impressions of possible impairment in FAS have not been investigated experimentally, as far as we know.

In our research we focus on the following questions: (i) whether FAS speech really sounds foreign as opposed to impaired, (ii) how the perceptual impressions of FAS speech compare to perceptual judgments of speech of genuine foreign speakers speaking English (L2), and (iii) what contribution vowel quality may have to the perceptual impressions of ‘accent’ / impairment in FAS speech.

Our FAS speaker was a 56 year-old native English right-handed male, who reportedly spoke with an Essex accent before his stroke and sounded to most people Italian (or Greek) after the stroke. We extracted 21 intonation phrases from a radio interview with this speaker and recorded the same material also from five other speakers: three L2 (Italian, Greek and French) speakers of English, and two Essex speakers as controls. The randomised stimuli from all speakers together were subjected to a perceptual test by ten native English listeners, who were asked to rate each stimulus on 1-4 scales for ‘foreign accent’ and ‘impediment’.

The analysis of the perceptual judgments yielded the following results. The agreement across listeners on the ratings of both foreignness and impediment was significant but weak (i.e. there was considerable inter-listener variability). The FAS speaker was perceived as foreign to a similar extent as the L2 speakers (and more so than the Greek L2 speaker) and as significantly different from both Essex speakers, who were judged ‘definitely English’. On the impediment scale the FAS speaker’s mean rating approached mild impediment. The Greek L2 speaker was judged to be on average more impaired than the FAS speaker, but the other L2 speakers and the control Essex speakers were judged to have no impairment. The accent most frequently attributed to the FAS speaker was ’Indian’ and Italian, with Spanish, ‘Mediterranean’, French and Greek also mentioned. The impressionistic phonetic analysis revealed a number of features in our FAS speaker’s speech that are commonly found in L2 speakers of (particularly) ‘Mediterranean’ languages: reduced aspiration of voiceless plosives, /h/ dropping, epenthetic vowels, lack of vowel reduction, some changes in vowel quantity and quality (especially of high front vowels and diphthongs), and perceived syllable-timed rhythm. The acoustic analysis of vowel quality overall confirmed the impressionistic phonetic analysis. Possible impact of the changes in vowel quality in the FAS speaker on perceptual judgments is discussed.

Acoustic analysis of prosody for Cantonese normal and aphasic discourse – A pilot study

Alice Lee[1], Anthony Kong[2], & Sampo Law[3]
Department of Speech and Hearing Sciences, University College Cork[1], Department of Communication Sciences and Disorders, University of Central Florida[2], Division of Speech and Hearing Sciences, The University of Hong Kong[3]

Background

Previous studies on prosodic problems in aphasia have focused the investigation on the sentence level. However, the evaluation should be based on connected speech as it is considered the most sensitive task to different dimensions of atypical prosody (Leuschel & Docherty, 1996). Recently, Tseng, Pin, Lee, Wang and Chen (2005) proposed a top-down, multi-layered framework, which considers all relevant levels (prosodic phrase group, breath group, prosodic phrase, prosodic word, and syllable) that constitute discourse prosody. While Tseng et al.'s model is developed based on speech samples of normal Mandarin speakers, the present paper reports a pilot study that has applied this framework to analyze discourse prosody in Cantonese-speaking normal speakers and individuals with aphasia.

Method

The subjects were two individuals with aphasia (one female of 47 years old with transcortical motor aphasia and one male aged 49 years old with anomic aphasia) and two age-, gender- and education level-matched normal individuals (one 44-year-old female and one 47-year-old male; education level was secondary or below). Each speaker partook in picture description (Cat Rescue picture from the AphasiaBank protocol) and story telling (Tortoise and Hare). The speech samples were recorded using a digital recorder and a condenser microphone in a quiet room. Acoustic-phonetic segmentation, boundary break annotation (based on Tseng et al.'s guidelines) and acoustic analysis were conducted using Praat version 5.1.02 (Boersma & Weenink, 1992-2009). The following parameters were measured: syllable duration, fundamental frequency (F0), and intensity. The data from the two tasks were combined for each parameter and standardised for each speaker. Linear regression was used for estimating the variance accounted for in duration, intensity, and F0 by each prosodic layer. The independent variables were: at the syllable layer – the type of consonant, vowel, and tone that each syllable contains, the type of consonant, vowel, and tone of the preceding and following syllable; the number of syllables and the position of the syllable in a prosodic word and a prosodic phrase, and the position of the prosodic phrase in a breath group (Tseng et al., 2005).

Results

Moderate interjudge agreement was found for boundary break annotation (range: 51%-77%). Overall, the factors at the syllable level accounted for less than half of the variance in each dependent variable and the inclusion of factors at the prosodic word, prosodic phrase, and breath group levels had largely reduced the residual errors by over 20%. The speaker with fluent aphasia showed less residual error than the two controls at each prosodic layer, while the speaker with non-fluent aphasia showed higher residual error when compared to the controls.

Discussion and conclusion

The pilot data suggests that the hierarchical prosodic phrase grouping framework and the linear models are able to give an overall picture of the discourse prosody of a speaker. It can be extended to capture the prosodic disturbance in aphasia through modification and additional parameters (e.g. measuring the use of inappropriate breaks based on language analysis).

Development of segment-specific coarticulation: an ultrasound study

Natalia Zharkova
Speech Science Research Centre, Queen Margaret University

Studying development of coarticulation contributes to understanding the emergence of the system of organised motor skills required for a given language. There are few developmental studies of coarticulation, partly because it is difficult to find instrumental techniques for imaging internal articulators in young children. With respect to lingual coarticulatory patterns, it is unclear how their nature and extent change with age. This study addressed this empirical gap by comparing lingual coarticulation in children and adults, using articulatory measures derived from ultrasound tongue imaging.

Evidence from articulatory phonetic studies of typical adult productions shows that lingual coarticulation induced on the consonant by adjacent vowels is dependent on the identity of the consonant. For example, alveolar and postalveolar fricatives have been shown across languages to exhibit different degrees of vowel-dependent coarticulation in consonant-vowel (CV) syllables: e.g. Fowler & Brancazio 2000; Recasens & Espinosa 2009. The cross-consonant difference has been explained by mechanico-inertial properties of the active articulator for the consonant. The current study investigated how cross-segment differences in coarticulation develop with age. Ultrasound data on consonant adaptation to the following vowel were compared in children and adults.

The participants, all native speakers of Standard Scottish English, were ten typically developing children aged five to nine years, and ten adults. The data were CV syllables including all CV combinations of the following segments: /s/, /ʃ/, /i/, /u/ and /a/. The syllables were produced in a carrier phrase. Synchronised ultrasound and acoustic data were collected. Ultrasound frames at the mid-point of the consonant were identified in each CV sequence. Extent of consonantal adaptation to the following vowel was compared in children and adults. For each consonant, the distance between the consonant contours in each pair of vowel environments was calculated, for each speaker individually, and then compared across age group and consonant.

In adults, consonant contours for /ʃ/ in different vowel environments were closer to each other than for /s/. In children the difference was in the same direction, but it was smaller. A significant interaction of age group and consonant was found, suggesting that while in adults the two fricatives are coarticulated in a noticeably different way, children have not yet fully developed this consonant-specific feature of coarticulation.

The results of this study are in agreement with the literature claims that the development of speech motor control continues through childhood and into adolescence (e.g. Walsh & Smith 2002). The segment-specific differences between children and adults reported in this study may be related to the maturation of fine control of the tongue dorsum for /ʃ/ and the tip/blade for /s/. Past literature showed that these two fricatives are differentiated to a smaller extent in children and adults, both articulatorily and perceptually (e.g. Nittrouer, Studdert-Kennedy & McGowan 1989). The current study provided evidence on how this greater differentiation in adults is manifested in coarticulation. Studies of more age groups are planned, in order to establish when children's lingual coarticulatory patterns become adult-like. In addition to measuring coarticulation based on individual time points during the CV sequence, dynamic information on tongue movements will be analysed and compared across age groups.

Phonetic settings in Belgian Standard Dutch: Evidence from electropalatography

Jo Verhoeven
City University London

Much of the phonetic description of speech has focused on individual speech segments of languages, often abstracting away from more general underlying pronunciation tendencies which may have an overall influence on the pronunciation of individual speech segments. Over the past decades, however, there has been increasing attention to the more overarching effects of such general pronunciation tendencies. A good example of this are phonetic settings. A phonetic setting is defined by Laver (1994) as "any co-ordinatory tendency underlying the production of the chain of segments in speech towards maintaining a particular configuration or state of the vocal apparatus" (p. 396). Laver makes a distinction between articulatory, phonatory and tension settings. The research reported here relates to articulatory settings, i.e. the assumed typical configuration of the vocal tract and the positioning of the articulators.

In the past, it has often been assumed that articulatory settings may vary between languages. Although this idea seems well accepted, it has proven to be particularly elusive to objective measurement. Some evidence has been provided in Gick et al (2004) on the basis of measurements of inter-utterance tongue rest position on X-ray films, while Lowie & Wander (2007) have observed differences in acoustic measurements of articulatory settings. The study that we would like to report has measured intra-speaker differences in articulatory settings by means of electropalatography.

The starting point of this investigation was the idea that Belgian students of drama have systematically been required in their drama education to use an alveolar trill realisation of /r/ instead of a uvular trill (both realisations are in free variation in Standard Belgian Dutch): students who naturally use a uvular trill have had to learn the alveolar trill. The underlying idea to this tradition has been that the use of an alveolar trill will create a shift to a more anterior articulatory setting which is assumed to lead to more intelligible speech which is important when speaking in the theatre. Similar ideas have existed in German drama education (Sieb: Bühnendeutsch).

The study presented here has investigated this potential shift of settings by means of electropalatography in one female speaker: this speaker is a Belgian television presenter/actress who naturally uses a uvular trill in everyday speech, while consistently using an alveolar trill in all her professional encounters. The speaker's command over both trill variants can be regarded equally good. This speaker was required to read a large number of short sentence pairs which differed minimally in the presence/absence of /r/: the sentences without /r/ were intended to provide information about the neutral setting, while the sentences containing /r/ were read on one occasion with a uvular trill and with an alveolar trill on another occasion, potentially tapping into different settings.

For each of the speaker's realisations a Centre of Gravity measure was obtained which provides information about the average position of tongue contact in the vocal tract. The results suggest that the phonetic realisation of the trill as either alveolar or uvular has a significant effect on the centre of gravity: alveolar trills lead to a more anterior setting, while uvular trills invoke a more posterior setting. The implications of this for the description of regional variants of Dutch will be discussed.

An Edinburgh Speech Production Facility

Alice Turk[1], Jim Scobbie[2], Christian Geng[1], Satsuki Nakai[1], & others
The University of Edinburgh[1], Queen Margaret University[2]

This EPSRC-funded facility is designed for the collection of synchronized articulatory and acoustic data during dialogue. It will be open to the international research community for funded use as of September 2010. As part of the EPSRC project, we recorded a corpus of dialogue sessions (8 are planned, 6 are recorded to date). The transcribed corpus will be available in 2010. We briefly describe the facility setup, and our recorded corpus.

1. The facility

The facility is built around two Carstens' AG500 electromagnetic articulographs (EMA). Synchronization of both EMA data sources and the acoustic waveforms is achieved by capturing (a) synch impulses of both machines and (b) the acoustic waveforms of both speakers by means of Articulate Instruments' Ltd hardware. This hardware includes an ADLINK DAQ-2213 8-channel, 16-bit differential input data acquisition card mounted to a standard PC and connected to an 8+4 Channel Analogue/Video Breakout Box. The hardware is also capable of synchronizing other time series data (EPG). Specific problems of the dual setup include a) electromagnetic inter-machine interference b) issues of temporal synchronization, c) the necessity of a talkback system, d) position estimation problems arising from e.g. larger amounts of head movement in dialogue as compared to monologue settings, and e) issues of position estimation arising from long trial durations. We will discuss these in a fuller presentation. Position-estimation procedures include those described in Hoole & Zierdt (2006) and unscented Kalman filtering-based algorithms, developed by K. Richmond. Data analysis software (Articulate Assistant Advanced, EMA module) has been commissioned from Articulate Instruments Ltd (2009).

2. The dialogue corpus

The articulatory corpus was designed with two main intersecting aims. The first was to elicit a range of speech styles or registers from the speakers, common in language use. The corpus includes several types of spontaneous speech and therefore provides an alternative to traditional reading corpora. The second was to extend the corpus beyond single-speaker monologue, by using tasks that promote natural discourse and interaction. A subsidiary driver was to use speakers from a wider dialect variety than is normally used. To this end, we recorded primarily Scottish English and Southern British English participants in dialogue with each other. Spontaneous dialogue was supplemented by other tasks, some undertaken in a preliminary acoustics-only screening session which included a modified and extended Wellsian Lexical set sample and a revised version of Comma Gets a Cure.

The articulatory data collection protocol includes a variety of tasks:

- Story reading (Comma Gets a Cure), lexical sets, spontaneous story telling, diadochokinetic tasks (monologue)
- Map tasks (Anderson et al. 1991), Spot the Difference picture tasks (Bradlow et al. 2007), story-recall (dialogue)
- Shadowing

Each dialogue session includes approximately 30 minutes of speech. We will exemplify the key characteristics of the corpus, presenting an articulatory perspective on the following phenomena: naturally occurring covert speech errors, accent accommodation, turn taking, and shadowing.

The corpus includes several types of annotations: 1) an orthographic transcription, 2) disfluencies, and 3) a modified ToBI transcription (for part of the corpus only). Further details and examples will be provided in the full presentation.

A corpus for cross-linguistic analysis of the phonetics of overlapping talk in talk-in-interaction

Emina Kurtić[1,2], Guy J. Brown[2], & Bill Wells[1]
Department of Human Communication Sciences[1], Department of Computer Science[2]
University of Sheffield

The phonetic and phonological properties of a language can be studied as a set of resources systematically deployed in the management of talk-in-interaction (Couper-Kuhlen & Selting 1996, Local 2007). This approach, often called interactional phonetics (Local 2007), relies on the availability of speech corpora that fulfil two criteria. Firstly, they consist of recordings of naturally-occurring speech (e.g. mundane conversation, meetings) rather than speech arising from activities that are designed specifically for the purposes of linguistic research. Secondly, the recorded speech signal must be of sufficiently high quality to allow for reliable automatic extraction of phonetic parameters (e.g. F0, intensity, spectral features). In the current state of technology, the latter is best achieved by recording each speaker on a separate microphone, preferably mounted on a headset. This approach is particularly important for studying overlapping speech. In the situation where more than one speaker talks, reliable estimation of phonetic parameters is only possible from the speech signal recorded on single channels, if signal processing approaches for sound separation which may introduce artefacts into the speech are to be avoided.

In this poster we describe a corpus of conversational speech in two languages, Bosnian Serbo-Croatian (BSC) and British English (BE), collected as part of an AHRC-funded project in progress. The corpus has been compiled to fulfil the requirements of interactional phonetic analysis in general and analysis of overlapping speech in particular. Three hours of conversation between four peers have been recorded in each language. Video recordings and audio recordings on separate channels have been made. We describe the procedure of data acquisition, its segmentation and transcription, as well as its storage and availability. Furthermore, we illustrate how phonetic and interactional properties of overlapping talk in both languages can be compared using this data.

Speaking in time and sequence: phonetics and the management of talk-in-interaction

John Local[1] & Gareth Walker[2]
University of York[1] University of Sheffield[2]

In trying to account for the phonetic variability and detail evident in a corpus of recorded speech, speech scientists typically attend to such factors as phonological and grammatical structure, speech style, emotional state and extra-linguistic social categories, among others. Beginning in the early 1980s, research into the phonetics of talk-in-interaction has shown that phonetic detail and variability can be shaped by interactional factors.

This poster presentation has two main objectives. The first is to set out the methodological imperatives which underpin our ongoing research into the phonetics of talk-in-interaction. The second is to present results arising from the most detailed study of turn transition — the basic social action of managing the change in speakership from one participant to another — to date. In the literature on the phonetics of talk-in-interaction it has been claimed that the completion of a tonic pitch movement marks the beginning of the transition space: the space at the possible end of a speaker's turn where a co-participant might legitimately start to talk. Based on detailed analysis of the phonetic and sequential details of almost 600 points of possible syntactic completion in unscripted everyday conversation we provide qualitative and quantitative evidence of the following:

1. Overwhelmingly, turn transition occurs where there has been a nuclear tone in that turn-constructional unit.
2. Turn transition does not generally occur where there has not been a nuclear tone in that turn-constructional unit.
3. Where a nuclear tone occurs but turn transition does not occur there are usually other aspects of the speech signal which project the imminent production of more talk from the current speaker. These features include: an absence of *diminuendo* and *rallentando* characteristics which regularly accompany other utterances which co-participants treat as complete; non-final resonance, anticipatory assimilation and final glottal closures at the end of the turn-constructional unit; the continuation of phonation into the talk which follows.

We provide various kinds of evidence that participants themselves orient to the phonetic features we describe as interactionally relevant, including the (non-)occurrence of turn transition at points of possible syntactic completion and the designing of incoming talk as turn-competitive in particular sequential environments.

This research is part of a larger project documenting the role of phonetic detail in the organisation of everyday interaction, with a particular focus on the management of turn transition, how speakers signal relationships between components of turns, and some of the ways in which a current turn can be shown to relate to some earlier (though not necessarily immediately prior) talk. This larger project will also be described. Audio extracts and speech analysis software will be made available during the poster presentation to provide for independent scrutiny, and discussion.

Prosodic matching of response tokens in conversational speech

Jan Gorisch
University of Sheffield, Department of Computer Science, Department of Human Communication Sciences

The phonetic forms and discourse functions of response tokens such as *mm* have been investigated for several decades. In general, previous studies have attempted to classify response tokens according to their communicative function, and then to search for prosodic similarities within each class (Gardner, 2001). Using data from naturally-occurring research meetings in the AMI corpus (native British and American speakers) we investigated the response token *uhu* in this way. Interactional analysis was used to identify instances of *uhu* that functioned either as continuers or as acknowledgments, and then we attempted to identify prosodic (and visual) cues which might differentiate these two functions. No difference in pitch patterns (F0 range, F0 movement) could be found between response tokens with the two different conversational functions: pitch characteristics could vary between extremes (e.g. rising and falling pitch movements) within classes that were distinct from an interactional point of view. Visual cues such as gestures and gaze did not predict these differences either. However, it was observed that pitch characteristics of tokens of the same class appeared to depend on the pitch characteristics of the immediate prior talk of the interactional partner. If, for example, the prior talk ended with rising intonation, the response token was also produced with a rise if it was encouraging the other speaker to continue talking. In order to perform the same action, the intonation of the “uhu” was falling if the previous talk ended with a fall. This was not the case when the utterer of the “uhu” was taking the floor or projecting to do so. In order to quantify these effects, a technique that objectively measures the similarity of prosodic features such as pitch movement and individual speakers' ranges is presented. Using some examples we explore how copying vs. non-copying behaviour in the prosodic domain of short response tokens is used to manage talk in multi-party interaction (Szczepek Reed, 2006). The analysis is based on the principal of cross-correlation where the similarity of two signals – here F0 contours – is established. The method is extensible to cover other prosodic cues such as intensity or tempo, and even visual cues such as head nodding.

Spot the different speaking styles: is ‘elicited’ clear speech reflective of clear speech produced with communicative intent?

Rachel Baker & Valerie Hazan
UCL Speech Hearing and Phonetic Sciences

Most studies investigating clear speech rely on speech that is elicited by instructing speakers to read materials clearly (e.g. ‘as if speaking to a person with a hearing loss’). Under such conditions, speakers produce speech which, compared to their casual speech, is enhanced in a number of acoustic-phonetic dimensions. However, does this clear speech have the same acoustic-phonetic characteristics as spontaneous speech produced in response to a real communication barrier?

We developed the “Diapix” task – a spot the difference task created by Van Engen *et al* (in press) – to collect spontaneous speech that contains similar words across different speakers and communicative conditions. For this task, two participants are seated in separate rooms and communicate over headsets to find twelve differences between two versions of the same picture. Differences relate to keywords belonging to minimal word pairs, e.g. “bear”/“pear”. Twelve diapix pictures that were matched for difficulty were created for the study.

Forty-four Southern British English speakers took part in the study, and worked in pairs. In the ‘casual speech’ condition, three diapix tasks were done in good listening conditions. In the ‘clear speech’ condition, six diapix tasks were done when one person heard the other via a three-channel vocoder, to simulate communication with a cochlear implant user. Each person took a turn at being the ‘impaired listener’ so that clear speech was collected from each speaker, i.e. when they were not experiencing the adverse listening condition. Additionally, each speaker read a set of sentences containing keywords that occur in the diapix tasks, both normally and when instructed to speak clearly.

Acoustic-phonetic analyses of the spontaneous speech show that compared to the casual speech, the clear speech had a slower speech rate, higher median f0 and higher mean energy between 1 and 3 kHz. This fits with previous findings regarding clear “read” speech and shows that it is possible to elicit clear speech through an interactive task without explicit instruction. To provide a more direct comparison of spontaneous and read clear speech, acoustic-phonetic characteristics of the casual and clear “read” speech will also be presented, to assess whether the change in these acoustic-phonetic characteristics are of the same magnitude in clear speech which is produced either with or without communicative intent.

A Cross-Language Study of Glottal Stop Perception

Yasuaki Shinohara
University College London

Introduction A single glottal stop is often produced between vowels in London Regional RP English (Wells, 1982; Fabricius, 2000; Cruttenden, 2008); while Japanese language does not contain the single intervocalic glottal stop as it precedes a plosive (Ito & Strange, 2009). Therefore, there may be perceptual differences for glottal stop between Londoners and Japanese and it is investigated in this research.

Method Subjects were 15 Londoners and 12 Japanese, and the stimuli were 'bear' and 'better' synthesized by Klattsyn. The vowel /e/ of 'bear' was controlled in the magnitude of critical acoustic cues of glottal stop: diplophonia (Di), fundamental frequency (Fx) dip and amplitude of voicing (AV) dip. 7 stimuli continua were composed in the parameters of Di, Fx, AV, Di&Fx, Di&AV, Fx&AV and Di&Fx&AV, and each continuum has 8 steps of the degree of the cues. Each step has +10 for Di, - 5 Hz for Fx and - 2 dB for AV. Consequently, the values of Di, Fx, AV for /e/ of 'bear' are 0 (Di), 110 Hz and 55 dB, and these of 'better' stimulus are 70 (Di), 75 Hz and 41 dB.

In the experiment, they firstly listened to the extreme stimuli 'bear' and 'better' played by approximately 80dB through a headphone in a sound booth; and they chose either of them by ticking according to which a stimulus was closer to, while listening to randomized 280 stimuli (8 steps×7 parameters×5 times each).

Phonetic boundaries between modal vowel and laryngealized vowel for /t/ and gradients of slopes on stimuli continua were compared between Londoners and Japanese by t-test of SPSS for investigating the perceptual differences.

Results Both language groups can perceive the acoustic cues signaling glottal stops. However, there is a significant perceptual difference between them in the place of the phonetic boundaries of the parameters: Di&AV (F=0.59, t=2.315, df=25, p<0.05) and Di&Fx&AV (F=0.136, t=2.249, df=25, p<0.05). This result suggests that Japanese can identify 'better' with relatively less Di&AV and less Di&Fx&AV.

The important acoustic cue for Japanese speakers to perceive 'better' seems AV dip. Since AV (p=0.052) and Fx&AV (p=0.080) continua also result in near significant difference, perceptual sensitivity of AV is higher for Japanese than for Londoners.

Discussion These perceptual differences may be caused by the different approaches to the task. Since English group often uses a single glottal stop for /t/ of 'better', they may perceive the glottal stop as linguistic function for /t/. On the other hand, Japanese speakers are not familiar with it so that they may complete the task by not linguistic but psychoacoustic approach. When they identify 'better', they seem to evaluate how much different the stimuli are from smooth voicing of 'bear' stimulus.

Conclusion The perceptual difference in AV drop is observed between Londoners and Japanese. The difference may be emerged from the different approaches to the task: linguistic approach for Londoners and psychoacoustic approach for Japanese.

Second-language experience and speech-in-noise recognition: the role of talker-listener accent similarity

Melanie Pinet, Paul Iverson and Mark Huckvale
UCL, Speech, Hearing and Phonetic Sciences

Previous work has demonstrated that there is an interaction between native (L1) and non-native (L2) accents in speech recognition in noise, with listeners being better at L1 or L2 accents that match their own speech. This study investigated how this talker-listener interaction is modulated by L2 experience and accent similarity. L1 southern British English (SE) and L1 French listeners with varying L2 English experience (inexperienced, FI; experienced, FE; and bilinguals, FB) were tested on the recognition of English sentences mixed in speech-shaped noise that was spoken with a range of accents (SE, FE, FI, Northern Irish and Korean-accented English). The results demonstrated that FI listeners were more accurate with strongly accented FI talkers, and were progressively worse for the other accents. The SE listeners, however, had a strong advantage for SE speech, but were similarly poor at understanding the other accents. Their recognition processes were thus selectively tuned to their own accent, rather than having the graded sensitivity of FI listeners. FE and FB listeners were more similar to SE listeners as their experience with English increased. In order to account for the listeners' accent recognition patterns, an accent similarity metric involving vowel measurements was applied to the talkers and listeners' speech (i.e., the subjects in the listening experiment were recorded reading the same sentences). The results demonstrated that there were significant correlations between speech-in-noise recognition and the acoustic similarity of the talkers' and listeners' accents. Overall, the results suggest that L2 experience affects talker-listener accent interactions, altering both the intelligibility of different accents and the selectivity of accent processing.

A comparative study of Japanese, Canadian English and French speakers' perception of acoustic correlates of speech contrasts

Izabelle Grenon
University of Victoria

Individuals learning a second language (L2) may experience difficulties in perceiving non-native sound contrasts. Some L2 sounds may be problematic even if they are used in the native language (L1) of the learners, and conversely, other sounds may not be problematic even if they are absent from the learners' L1. Accordingly, a simple comparison of phonemic inventories between one's first and target language is insufficient to predict the impediments encountered by adult L2 learners. Models such as PAM (Best 1995) and SLM (e.g. Flege 2005) propose to look at cross-language perceptual similarity of speech sounds to make predictions about L2 perception or acquisition. However, perceptual similarity between L1 and L2 sounds cannot be predicted; it must be measured empirically (Bohn 2005), and consequently cannot readily be generalized to perception of novel contrasts or other languages. In line with experiments such as Morrison (2002), this research proposes an alternative way of looking at L2 difficulties. It is argued that perception of L2 contrasts may be best evaluated by investigating L2 listeners' sensitivity to different acoustic cues characterizing L2 sounds when making L2 judgments (rather than L1 judgments as for PAM or SLM).

The same acoustic cue may be used cross-linguistically for different phonological contrasts. For instance, Japanese speakers use vowel duration to contrast short and long vowels (e.g. Akamatsu 1997), while English speakers use vowel duration to distinguish coda consonants, or stressed and unstressed syllables (e.g. Ladefoged 2001). Crucially, it is questionable whether L2 learners may be able to capitalize on their sensitivity to acoustic cues used in their L1 to perceive a new phonological contrast using the same cues, even if the L2 phonological contrasts are not contrastive in the L1.

This study was designed to assess native and non-native speakers' perception of acoustic cues characterizing two English contrasts, reporting response time and identification judgments in a forced-choice task using manipulated natural speech samples. Experiment 1 evaluates the use of vowel quality (F1/F2 ratio) and vowel duration in the perception of the “beat-bit” contrast by native Canadian English, Canadian French and Japanese speakers (N = 24 per group). Experiment 2 evaluates the use of vowel duration and voicing (glottal pulse) during the closure of the final consonant in the perception of the “bit-bid” contrast by the same groups of speakers.

Results of these experiments suggest that even if an acoustic cue is used differently in their L1, language learners are able to capitalize on their sensitivity to this cue to perceive non-native contrasts. Conversely, if this cue is generally ignored in their L1, listeners may rely on another cue instead. These results have important implications for language teaching and for further studies of L2 speech perception by showing that L2 learners are able to make, in some cases, near native-like L2 judgments based on the appropriate contrastive acoustic cues, even when the L2 sounds are generally assimilated to the same L1 category. Provided that sensitivity to specific acoustic cues may generalize to other contrasts, this method offers a potentially more effective and reliable way of evaluating the exact causes of the impediments encountered by L2 learners.

A Phonologically Calibrated Acoustic Dissimilarity Measure

Greg Kochanski, Ladan Baghai-Ravary & John Coleman
University of Oxford Phonetics Lab

One of the most basic comparisons between objects is to ask how similar they are. Linguistics and phonology are founded on this question. The classic definitions of phonemes and features involve contrast between minimal pairs. A minimal pair of words requires that there be two sounds that are dissimilar enough for the words to be considered different. Otherwise we wouldn't speak of a minimal pair of words but rather of a single word with two meanings.

Likewise, phonetic similarity is needed to group together separate instances into a single word or sound. Without some intuition about which sounds are so similar that they should be treated as instances of the same linguistic object, the field would be no more than a collection of trillions of disconnected examples.

So, it is important to have a measure of dissimilarity between sounds. Some exist already: e.g. measures of dissimilarity between the input and output of speech codecs have been used as a way of quantifying their performance (e.g. Gray, Gray, and Masuyama 1980). Cepstral distance has been frequently used, and the Itakura-Saito divergence is also widely used.

But, none of these have been explicitly calibrated against the differences in speech that are important to human language. In this paper, we do so.

As our test data, we collected a corpus of speech where many short phrases were recorded several times each. We paired the phrases and divided them into two classes: where the pairs were recorded from the same text *vs.* pairs from different texts. We then computed distances within all the pairs via a dynamic time-warping algorithm and constructed two histograms: same text and different text. From these histograms, we computed a numerical measure of how much they overlapped each other (it is effectively a t-statistic).

We can then compare different algorithms and/or variations on algorithms. We explored variants on the Itakura-Saito distance with an parametrized pre-filter, and also a Euclidean distance computed on an approximation of the perceptual spectrum. In the latter, we included parameters to individually scale all the components of the vector.

Within each algorithm, we used a simulated annealing algorithm to vary the parameters and find the minimum overlap. We found that the maximization made a dramatic difference in both cases. In both cases, the two histograms started out strongly overlapped, so that distances between different texts were often smaller than distances between utterances recorded from identical texts. However, after the maximization, the histograms were well separated: large distances were reliably associated with different texts, and vice versa.

We showed that an optimization procedure can tune a measurement of acoustic distance so that it corresponds well with the linguistic same-text/different-text dichotomy. We suggest that this technique can be valuable for quantifying similarity and dissimilarity in phonetics and phonology.

Speech and language therapy (SLT) students' production and perception of cardinal vowels: a case study of 6 SLT students

Jussi Wikström
University of Reading

The present study investigates the problems faced by six anglophone Speech and Language Therapy students producing and perceiving cardinal vowels around the time of their final practical phonetics examination, the extent to which the same sounds and features of sounds are difficult in both production and perception, and the extent to which the participants' confidence ratings are reflected in their performance. It also looks at how well the participants retained the relevant phonetic knowledge over a seven-month period after their practical phonetics examination. The participants' productions of cardinal vowels were recorded and compared with those of the lecturers who had been involved in teaching their practical phonetics module. The participants did a perception test and completed a questionnaire focussing on their confidence with regards to producing and perceiving the different cardinal vowels. It was found that cardinals 2, 4, 7, 11, 12, 15 and 16 were particularly difficult to produce and cardinals 5, 9, 12, 14 and 15 were the most challenging to perceive. Although the participants' confidence ratings reflected their performance well overall, they appeared to be overly confident about their ability to produce cardinals 2 and 10 and their ability to perceive cardinals 13 and 16 while their confidence ratings suggest they were under-confident about their ability to produce cardinals 10 and 12 and about their ability to perceive cardinals 3, 6 and 10. The participants retained the phonetic ability necessary to produce and transcribe cardinal vowels over the 7-month period after their final practical phonetics examination, performing only slightly worse in the second production test and somewhat better in the second perception test overall, but one participant performed much worse in both tests in the second testing session, suggesting that there may be significant variation between individual students.

The role of short-term memory in phonetic transcription

Esther Maguire and Rachael-Anne Knight
City University London

Transcription is a vital skill for any phonetician. Knowing more about the skills and abilities used during phonetic transcription would improve our understanding of the process, and allow us to better facilitate teaching and learning. This study focused on the potential role of short-term memory in transcription.

It was hypothesised that short-term may play an important role during the phonetic transcription of verbally presented material. Phonological short-term memory, (e.g. Baddeley & Hitch, 1974) involves holding verbally presented material in a phonological loop where it can be mentally rehearsed. Knight (in press) found improvements to the accuracy of nonsense word transcriptions when more repetitions were presented. It was suggested that this might be due to the nonsense words exceeding the limited capacity of the phonological loop, and that extra repetitions are beneficial as they allow a different section of the signal to be stored in the loop each time.

In this study the short-term memory abilities of 16 Speech and Language Therapy undergraduates were assessed using nonword repetition and digit span tasks. These results were correlated with the subtests of the students' university assessments in phonetics. Significant positive correlations were found between students' digit span and assessment results in phonemic, allophonic and nonsense word transcription. As expected, no significant correlations were found between short-term memory and substitution tasks, in which only a small amount of material must be remembered. In addition, no significant correlations were found between nonword repetition and any of the phonetic subtests.

These results suggest that some component of memory does indeed play a role in phonetic transcription tasks that require attention to longer passages of speech. However, the different findings for nonword repetition and digit span tasks suggest that it is not simply a case of using short-term memory to store the material that is presented, but of using *working* memory to manipulate that material whilst it is stored in the phonological loop. These results are used to inform a cognitive model of phonetic transcription, and implications for teaching and learning are considered.

Young and old listeners' perception of English-accented speech in a background of English- and foreign-accented babble

Antje Heinrich, Korlin Bruhn &, Sarah Hawkins
Department of Linguistics, University of Cambridge

Foreign–accented speech is typically more difficult for native speakers to understand than speech uttered by a native speaker, particularly when heard in a noisy background (e.g., Tajima, Port, & Dalby, 1997). However, it is unknown whether foreign accents are similarly disadvantageous for intelligibility when they occur in the ignored background voices rather than in the attended target speech. Experimental evidence that background speech is easier to ignore when in a foreign language that cannot be understood (Van Engen & Bradlow, 2007) suggests that foreign-accented English in the background might also be easier for native speakers of English to ignore than native-accented English.

The current study measured identification accuracy of final target words in English sentences read by a native speaker of Southern Standard British English and heard in background (English) speech babble. Variables were target word predictability, number of voices contributing to the background speech babble, and native accent of the babble. Predictability was varied by changing the words preceding the target word in phonetically-controlled ways so that the target was highly predictable or highly unpredictable as rated by a number of judges. The masking speech babble contained one, three, or eight talkers of the same accent reading from an English text of their choice. The accents of the babble were British English, American English, Indian English, and Italian English. 40 young and 40 old listeners, all native speakers of British English participated in the perceptual experiment. In a between-groups design, each listener heard each target word only once, and in just one language accent in the babble, but with all three numbers of talkers (one, three and eight).

Preliminary analyses for young adults show main effects of babble accent, talker number, target word predictability, and an interaction between babble accent and talker number. Target words that were predictable from the preceding sentence were easier to understand than target words with low predictability from context. One competing talker was easier to ignore than three or eight talkers, and Indian English was more effective as a masker than any of the other accents, especially, in the 8-talker condition. We are currently testing older listeners on the same task. Interpretation will be in terms of spectral and phonetic properties of the masking babble relative to the signal.

Investigating the time course of perceptual adaptation to unfamiliar accented speech

Bronwen G. Evans & Emily C. Taylor
Speech, Hearing and Phonetic Sciences, University College London

Recent work in speech perception has demonstrated that although listeners are initially slower to recognize words spoken by a foreign-accented speaker (e.g., Adank et al., 2009), they are able to adapt after only a small amount of exposure (Clarke & Garrett, 2004). However, the evidence from studies of adaptation to regional accented speech is mixed. Although some studies have shown that listeners are able to adapt after only a short amount of exposure (e.g., Floccia et al., 2006), others have shown that listeners do not always alter their perceptual representations when listening to a non-native regional accent even if they are highly familiar with that accent (Evans & Iverson, 2004, 2007).

In this study, we further explored perceptual adaptation to different accents by comparing the time course of adaptation to an unfamiliar regional accent with adaptation to foreign-accented speech. Southern English listeners identified sentences in noise produced in either an unfamiliar regional accent (Glaswegian: GE) or an unfamiliar foreign-accent (Spanish-accented English: SpE). A baseline condition (native accent; SSBE) was also included in order to probe whether adaptation to an unfamiliar accent differed from adaptation to an unfamiliar talker from a familiar accent background. The results demonstrated that listeners were faster to adapt to their native accent, than either GE or SpE. Adaptation patterns for unfamiliar regional vs. foreign-accented speech also differed. Although listeners showed greater adaptation to foreign-accented speech they performed more poorly with foreign-accented speech overall and the rate of adaptation was slower. Overall, the results provide additional evidence that listeners are selectively tuned for their own accent (Preece-Pinet & Iverson, 2009), and suggest that although listeners are able to tune-in to an unfamiliar accent to a certain extent, learning continues over a longer period of time and is not as complete as at first thought.

Accent distribution in Punjabi English intonation

Kyle Day & Sam Hellmuth
University of York

Intonation can cause problems in understanding between L1 and L2 speakers of the same language. This study examined how the L1 Punjabi language influences L2 English in the emerging dialect of Punjabi English, and which L1 features speakers retained in i) rhoticity, ii) nuclear contour types (cf. Grabe et al 2005) and iii) distribution of pitch accents.

A corpus of read speech utterances in Bradford Punjabi English drawn from the IViE corpus (http://www.phon.ox.ac.uk/IViE) was submitted to prosodic transcription and auditory analysis, alongside a parallel set of utterances from the IViE corpus from speakers of Leeds English, which provided a control group to compare the two dialects. Transcription was carried out by the first author, with a subset of the data also transcribed by the second author in an inter-transcriber agreement study.

There was little variation in the realisation of rhotics in the Punjabi English data, with the speakers using alveolar approximants in prevocalic position and showing no post-vocalic rhoticity. This means that we can classify the IViE corpus speakers as similar to the speakers who identified themselves as 'British' (rather than 'British Asian') in the categorisation of Hirson et al 2007. We might therefore expect the prosodic aspects of their speech to be also be somewhat more English-like than Punjabi-like.

In fact, however, our survey suggests that Punjabi English intonation differs from (nearby) Leeds English intonation in two ways. Firstly, there appears to be somewhat more variation in the choice of nuclear and pre-nuclear contours in the Punjabi English data than in the Leeds English data. An even stronger distinction is seen in the distribution of pitch accents between the two sets of speakers. The Punjabi English speakers produced a significantly greater number of pitch accents than Leeds English speakers, and the accent distribution patterns within individual utterances also varied between the two dialects. Both of these properties have been found to also hold of Tamil English and Gujerati English (Wiltshire & Harnsberger 2006).

Given the parallel with the prosodic patterns of other Asian Englishes, and the fact that these are known to resist de-accenting (Cruttenden 2006, Ladd 2008), we also report the results of a small additional survey of the intonational properties of utterances containing a contrastive focus, extracted from spontaneous speech Map Task data. We found that deaccenting was not prevalent in the spontaneous speech, and that pitch accent distribution is as high in spontaneous speech in Punjabi English as it is in read speech. We interpret these findings as evidence of a greater tendency to suprasegmental transfer than segmental transfer in Punjabi English (cf. Anderson-Hsieh et al 1992, Munro 1995).

Accent attribution in speakers with Foreign Accent Syndrome

Jo Verhoeven[1], Peter Mariën[2], Michèle Pettinato[1], & Guy De Pauw[3]
City University London[1], Vrije Universiteit Brussel[2], Universiteit Antwerpen[3]

Foreign Accent Syndrome (FAS) has traditionally been defined as an acquired motor speech disorder in which the speech of a patient is recognised as a foreign accent by members of the same speech community as the patient (Whitaker, 1984). The disorder has been documented in at least 81 patients since it was first anecdotally described by Marie in 1907 (Marie, 1907). Although it is the only motor speech disorder which is defined in terms of the perceptual impression it invokes in listeners, the systematic perceptual investigation of foreign accentedness in FAS speakers itself has hardly received any attention at all (Di Dio et al 2006).

In the study that we would like to report, a perception experiment was carried out in which three groups of listeners attributed accents to three groups of speakers. Speaker group 1 consisted of 5 speakers of Belgian Dutch who had previously been diagnosed with FAS; speaker group 2 consisted of 5 non-native speakers of Dutch with a real foreign accent and speaker group 3 consisted of 5 native speaker controls without any trace of a foreign accent, but with a clearly identifiable regional Dutch accent. Samples of spontaneous were collected from the speakers and these speech samples were presented to three listening panels for accent attribution. These listening panels differed in their degree of familiarity with foreign accents: one group (n = 42) had no formal experience with rating foreign accents (University students in Psychology), one group (n = 37) was highly experienced in assessing speech and language pathology (Speech and Language Therapists), while the third group (n= 44) was highly familiar with a wide variety of foreign accents (Teachers of Dutch as a Foreign Language). Besides having to attribute an accent to the speakers, the listeners also rated their own confidence in their accent ratings.

The nominal data obtained in this experiment were analyzed by means of a correspondence analysis, while the numerical data (confidence judgements and interrater agreement scores) were analysed by means of analysis of variance. The results strongly indicate that the speakers with FAS are indeed often attributed a foreign accent, but that they clearly differ from speakers with a real foreign accent on all the dimensions that were analyzed. FAS speakers are most strongly associated with a (from a Belgian perspective) more familiar foreign accent such as French, while the real foreign speech samples are fairly consistently associated with more exotic accents such as e.g. Eastern European. In addition, the FAS speakers are relatively often recognized as native speakers of Dutch, while the real foreign accents hardly ever are. Furthermore, listeners were generally less accurate in identifying the true linguistic background of the FAS speakers. In addition, the listening panel was less certain about their accent attributions and less consistent in FAS as compared to speakers with a real foreign accent.

Taken together, these results suggest that the foreign accent associated with FAS is not entirely in the ear of the beholder, but that there are certain pronunciation characteristics in FAS which listeners are willing to interpret as foreign. This foreignness is more of a ‘generic’ than of a ‘specific’ kind. It is hypothesized that FAS foreignness in listeners is triggered by very specific changes in the pronunciation characteristics of patients due to more familiar motor speech disorders such as dysarthria and/or apraxia of speech. Unaffected regional speech characteristics of FAS speakers may subsequently reinforce the perception of a specific foreign accent.

Acoustic Cues to Givenness in Foreign Accent Syndrome

Anja Kuschmann[1], Anja Lowit[1], Nick Miller[2] and Ineke Mennen[3]
Speech and Language Therapy Division, Strathclyde University[1]; Speech and Language Sciences[2], Newcastle University; ESRC Centre for Research on Bilingualism, Bangor University[3]

Research on healthy speech has proposed that stressed words or syllables are in general longer, louder and higher in pitch than unstressed words, with pitch being the primary perceptual cue to prominence in stress-accent languages (Beckman 1986). There is evidence that the respective acoustic parameters involved in conveying the relevant prosodic information - i.e. duration, intensity and fundamental frequency (F0) - can be subject to cue trading relations (Cooper et al 1985). This phenomenon is thought to be of particular relevance in disordered speech where individuals who experience difficulties adjusting acoustic parameters may compensate by relying more heavily on parameters they still have control over (Patel & Campellone 2009). In Foreign Accent Syndrome - a speech disorder, where acoustic changes lead to a perceived foreign accent in speech - difficulties with word stress have been reported (e.g. Blumstein & Kurowski 2006). However, little is known about the use of the relevant acoustic parameters and their impact on the ability to convey prosodic-linguistic information.

The present study reports the results of a production experiment that examined the use of acoustic cues to mark new and given referents within short sentences by four speakers with FAS and four healthy control speakers.

Disyllabic trochaic target words were embedded in carrier sentences, and controlled for givenness status (new vs. given) and sentence position (initial vs. medial vs. final). A *praat* script was employed to extract the length of the stressed syllables of the target words, the peak intensity on these intervals, and Hz values at specified points within the sentence. Within-speaker analyses were conducted for each acoustic parameter using a series of two-factor repeated measures ANOVA. For between-speaker analyses the data were normalised using z-transformation, and percentage differences were calculated.

Results show that speakers successfully differentiated between new and given elements by manipulating the different acoustic cues. The target syllables of new referents were significantly longer and higher in intensity and F0 than those of given referents. Although both groups employed all three acoustic parameters, differences across speakers were observed as to the combination of cues used. While in control speakers the relevance of the durational cue decreased over the course of the sentence, most speakers with FAS showed the inverse pattern. This outcome suggests that these speakers rely more heavily on durational cues than the control speakers.

Colour and luminance associations with vowel sounds in synaesthetes and non-synaesthetes

Anja Moos, David Simmons & Rachel Smith
University of Glasgow

About 4.4% of the population (Simner et al. 2006) have a neurological condition called synaesthesia: a cross-wiring of the senses where stimulation in one sensory modality leads to an automatic involuntary additional other sensation. Several types of synaesthesia are triggered by linguistic units: of these, grapheme → colour synaesthesia is the most common (here, graphemes are inherently coloured); phoneme (or phone) → colour synaesthesia is reported, but poorly understood. Colour associations are constant in a given synaesthete, but whilst there can be large individual differences, there is also a surprising level of consensus between synaesthetes (Simner 2007).

Previous studies have found systematic relationships between colours and vowels. Wrembel (2007) found that Polish students have similar patterns in associating English and Polish phonemes to colours. Marks (1975) found in a meta-analysis that the ratio of F2/F1 correlates with the green-red colour scale, such that green was associated with 'diffuse' vowels and red with 'compact' vowels. Many studies (e.g. Jakobson 1962) have suggested a correlation between luminance and articulation place in which front vowels are usually perceived as light and back vowels as dark with Marks suggesting that luminance correlates with F2 which is supposed to represent the 'pitch of the vowels'.

The aim of our study was therefore twofold: (1) Are there systematic relationships between vowel sounds and colours? (2) Do these relationships differ between synaesthetes and non-synaesthetes?

We have systematically measured the association between vowel sound and brightness/colour in synaesthetes and the general population. Participants had to choose one out of 16 colours in the first part of the experiment and grey shades in the second part while listening to one out of 16 vowel sounds. The eight primary cardinal vowels were recorded by a male phonetician and eight intermediate vowels were synthesised using a morphing procedure (Kawahara 2006). Each vowel was presented 16 times in a randomised order to test for consistency. Of the three grapheme → colour synaesthetes we tested, two gave random responses to vowel-grey shade associations, and one perceived front vowels as lighter than back vowels. Two synaesthetes confirmed Marks hypothesis and associated green with front vowels, reddish with open vowels and darker colours with back vowels. In one synaesthete we found that position on the vowel quadrilateral of her colour associations correlated with position on the standard hue circle. Results for the 20 non-synaesthetes, analysed using general linear mixed-effects modelling, showed low within-subject consistency, but significant correlations insofar as front vowels were associated with light luminance and back vowels with darker luminance, F1 having a stronger influence on the choice of luminance than F2. It also showed a relationship between open vowels and red (high F1 and small F2/F1 ratio) and front close vowels and green (big F2/F1 ratio).

Association patterns are similar in synaesthetes and non-synaesthetes but there is a big difference in consistency of the vowel-colour associations. Both groups seem to use similar mechanisms for their vowel colour associations regarding acoustic measures and articulation place on one side and colour measures on the other. A remaining question is whether cross-modal perception is common to us all but variable in vividness or whether there is a clear qualitative division between synaesthetes and non-synaesthetes.

Recordings of disordered speech: The effect of 3 different modes of audiovisual presentation on transcription agreement.

Tom Starr-Marshall.
Talkativity & City University London.

This study explores the possible advantages and/or disadvantages of using audio/video recordings in making phonetic transcriptions of disordered speech. The reliability of phonetic transcription in the assessment of children with speech disorders is essential in Speech and Language Therapy for the purpose of accurate differential diagnosis and the appropriate selection of intervention targets and approaches. However, there are currently no Royal College of Speech and Language Therapists (RCSLT) guidelines or position statements on standard procedures for assessment using phonetic transcription for children with speech disorders except that some assessment of phonology and articulation should take place. Many research papers have used recording in their procedures but few have specifically targeted gaining empirical data on the effect different conditions of audio/video recording may have on the reliability of phonetic transcription. Two main issues have been postulated in the literature that may have major influences on the reliability of recorded data. Firstly, whether visual information aids transcription by giving visual cues to the place of articulation of sounds articulated at the front of the mouth or whether certain visual based psycho-acoustic effects such as the McGurk Effect (McGurk et al., 1976) could give a false percept of place of articulation. The second issue first postulated by Warren and Gregory (1958) and further discussed by MacKay et al. (1993), is whether the perception of acoustic patterns changes the more times you hear them, this is known as the Verbal Transformation Effect. If a Verbal Transformation effect were to occur one might expect that listening to multiple presentations of a sound may cause the listener to normalize their judgements of the disordered sounds articulated. With the aim of creating an evidence base, from which to produce standard procedures for the use of recording in making transcriptions of disordered speech, this study attempts to answer the following questions:
1. Is the agreement of transcriptions of disordered speech increased or decreased by utilizing visual information regarding a speaker's lip movements?
2. Is the agreement of transcriptions of disordered speech increased or decreased by listening to disordered speech spoken information several times?
3. What effect does the severity of a speech disorder have on the agreement of transcriptions in both conditions of audio/visual presentation?
To answer these questions transcriber agreement was calculated across 3 modes of audiovisual presentation of recordings of children with severe and mild speech disorders
The conditions were as follows:

1. Single presentations of video clips of a child’s response for each of the 10 words in the Diagnostic Evaluation of Articulation and Phonology (DEAP) screening section (to mimic live conditions)
2. 5 x presentations of video clips of a child’s response for each of the 10 words in the DEAP screening section.
3. 5 x presentations of just the audio from video clips of a child’s response for each of the 10 words in the DEAP screening section.

Analysis of the data found that severity of speech disorder significantly lowered transcriber agreement but all other effects were insignificant. This study, therefore, concludes that for mild speech disorders clinicians may use the method they find easiest but perhaps for severe speech disorders more novel methods of assessment are needed.

PLENARY

Speech on the brain - the cortical processing of spoken language.

Sophie Scott
Institute of Cognitive Neuroscience, UCL

The aim of the presentation is to explore how functional and anatomical imaging can shed light on aspects of human speech production and perception. I will address the potential role of hemispheric asymmetries in speech perception, and also the ways that different kinds of anatomical and functional streams of processing are recruited for different aspects of speech perception. I will show how auditory areas are strongly involved in speech production. I will address the issue of expertise in speech perception by using functional and anatomical scans of trained phoneticians and professional voice artists. I will show how the data from the phoneticians shows certain brain areas which are correlated in size with the number of years transcription experience, and also some preliminary data suggesting a cortical differences which are not associated with training, and which may point to an existing predisposition. I will finish with some data showing fundamental differences in the patterns of cortical activation that we see when professional voice artists change their voices to command.

ABSTRACTS

Tuesday 30th March 2010

Durational evidence for word-based vs. Abercrombian foot constituent structure in limerick speech.

Alice Turk[1], & Stefanie Shattuck-Hufnagel[2],
University of Edinburgh[1], MIT[2]

A growing body of evidence suggests that a hierarchy of word-based constituents influences the phonetic shape of utterances. These constituents include word-sized and larger constituents, e.g. prosodic words, clitic groups, phonological phrases, and intonational phrases. Previous experiments have shown that word-rhyme durations in phrasally-stressed words depend on the number of syllables in the word (e.g. *–un* in *tuna* shorter than *–un* in *tune*), and possibly larger units (Huggins 1975, Beckman & Edwards 1990, Turk & Shattuck-Hufnagel 2000). Polysyllabic shortening is one of the mechanisms proposed to account for shorter durations in constituents with more syllables.

In this paper, we ask whether polysyllabic shortening evidence supports another type of proposed prosodic constituent: the cross-word, or Abercrombian foot (Abercrombie 1965). Abercrombian feet consist of a phrasally prominent syllable followed by non-phrasally stressed syllables up to, but not including the following phrasally prominent syllable. Table I shows that for some word sequences containing phrasal prominences on both content words, word-based constituents and Abercrombian feet are isomorphic. However, for e.g. *bake elixirs* and *bake avocadoes*, Abercrombian feet, but not word-based constituents, include word fragments.

Word-based constituents	Abercrombian Feet
[Bake] [apples	[Bake] [apples
[Baking] [apples	[Baking] [apples
[Bake us] [apples	[Bake us] [apples
[Bake] [avocadoes	[Bake avo-] [-cadoes
[Bake] [elixirs	[Bake e-] [-lixirs

Our materials consisted of phrases like those listed in Table 1, created out of 10 monosyllabic verb stems (*–bake, pick, cook, tab, bag, stop, track, grab, crib, catch)*. Our materials were recorded by six speakers of a variety of American English. We embedded each phrase in the 4th line of a limerick, in order to ensure reliable placement of phrasal prominences on each content word in the target sequence, and to encourage the production of near-isochronous inter-stress intervals.

There once was a boy from St. Paul; *Who loved to bake fruit in the fall*; *He'd give up his Snapple*; *To bake apples*; *With butter and sugar and all.*

We predicted that if polysyllabic shortening occurs within Abercrombian feet, e.g. *–ake* in *bake avocadoes* and *bake elixirs* should be shorter than *–ake* in *bake apples.* On the other hand, if polysyllabic shortening occurs within word-based constituents, but not Abercrombian feet, we expected e.g. shorter *–ake* in *baking apples* than in *bake apples*, but no difference between e.g. *–ake* in *bake apples* and *bake avocadoes*.

Preliminary results for three speakers suggest that word-based constituents are more influential than Abercrombian feet, even in limerick speech, which is highly encouraging of rhythmicity. In particular, we found that 1) inter-stress intervals are not isochronous: interval durations increase with increasing number of syllables; 2) word-based constituents influenced duration patterns for all three speakers; 3) speakers inserted other types of boundary markers (e.g. silence, glottalization) at word boundaries within Abercrombian feet, and 4) only one of the three speakers showed evidence of polysyllabic shortening within Abercrombian feet. The magnitude of shortening within this constituent was less than what we observed within word-based constituents, but suggests that speakers may have the option of using both word-based and Abercrombian feet simultaneously.

How far can phonological properties explain rhythm measures?

Elinor Keane[1], Anastassia Loukina[1], Greg Kochanski[1], Burton Rosner[1] & Chilin Shih[2]
Oxford University Phonetics Laboratory[1], EALC/Linguistics, University of Illinois[2],

Speech rhythm has long been thought to reflect the phonological structure of a language (e.g., Roach 1982; Dauer 1983, 1987). Syllable structure is a key example: languages that allow complex consonant clusters would have a rhythm characterized by much more variability in consonant length than a language like Mandarin where consonant clusters are rare. We explored this idea experimentally by seeing how well a range of popular rhythm measures were predicted by the phonological properties of the text.

The results are based on 3059 paragraphs read by 62 native speakers of English, Greek, French, Russian and Mandarin. The paragraphs were selected from the novel *Harry Potter and the Chamber of Secrets*, to represent the full range of phonological variation existing in each language. They included pairs of paragraphs chosen for particularly high and particularly low values of eleven different phonological properties. These were calculated from the expected transcription and included the average complexity of consonant clusters, percentage of diphthongs in the text and average sonority (assigning a sonority level of 0 to obstruents, 1 to sonorants and 2 to vowels).

First, we confirmed that languages indeed have different phonotactics, based on the expected transcription. For example, the complexity of consonant clusters in the English data was significantly greater than in the Mandarin data. A classifier based on a pair of averaged phonological properties (e.g. mean consonant cluster length and mean sonority) would correctly identify the language of 70% to 87% of the paragraphs (1Q-3Q range, depending on the pair of properties, chance=20%).

The recorded speech was divided into vowel-like and consonant-like segments using a language-independent automatic segmenter, trained on all five languages. From this, we computed 15 statistical indices proposed as rhythm measures in the literature, e.g. %V, VnPVI (references in Loukina et al. 2009): all were devised to capture durational variability between languages. In contrast to the classifiers based on phonological properties, we found large overlap between languages.

Phonological properties were found to predict paragraph-to-paragraph differences in rhythm measures rather poorly. The largest correlations involved the percentage of vowel-like segments (including sonorants) in speech vs. the percentage of voiced segments in text, but these only explained 9% of the variance in Russian and 18% in Mandarin. Instead, interspeaker differences accounted for much more of the variation in the rhythm measures in a linear regression analysis. For example, for Russian, the average adjusted r^2 across different rhythm measures was .112 for regressions against phonological properties, but .295 for regressions against speakers. The corresponding values for English were .139 and .335.

These results indicate that differences in timing strategies between speakers, even within the same language, are at least twice as important as the average phonological properties of the paragraph. It suggests that rhythm, in the sense of durational variability, may be determined more by performance differences between individuals than differences in the phonological structure of languages.

Speech rhythm: the language-specific integration of pitch and duration

Ruth Cumming
Department of Linguistics, University of Cambridge

In an attempt to quantify languages' rhythm, 'rhythm metrics' have become widely-used in phonetic experiments on rhythm, e.g. Ramus et al. (1999) (*%V, ΔV, ΔC*), Grabe & Low (2002) (*PVI*), Dellwo (2006) (*VarcoV*), Bertinetto & Bertini (2008) (*CCI*). Such metrics are based on statistics of the variability of vowel and consonant durations. However, in their current form, these metrics have several flaws: (i) speech production data only demonstrate how rhythm is spoken, even though it was purely auditory observations that first gave rise to the notion that languages *sound* rhythmically different (e.g. Steele, 1775; Classe, 1939), suggesting rhythm is worth examining from a listener's perspective (cf. Kohler, 2009); (ii) duration measurements mainly indicate timing, even though rhythm involves the acoustic *nature* (as well as timing) of prominences, so various other cues are worth investigating, e.g. f0, amplitude, spectral properties (cf. Lee & Todd, 2004); (iii) universally applied metrics ignore the possibility that rhythm may be perceived differently by speakers of rhythmically different languages, despite evidence that native language influences listeners' perception of segmental contrasts (e.g. Strange, 1995) and prominence (e.g. Dupoux et al., 2008).

The experiment reported here addressed these problems, and tried to nudge rhythm research out of the rut where it has become stuck (cf. Nolan & Asu, 2009). In a perceptual task, the stimuli were sentences constructed with a specific prosodic structure, in which f0 and duration were systematically manipulated on one prominent syllable. In each trial, listeners heard 9 stimuli (lexically identical sentences, but 3 f0 manipulation conditions x 3 duration manipulation conditions), and judged which sentence had the most natural rhythm. Three listener groups completed the task in their native language: Swiss German (SG), Swiss French (SFr) and (metropolitan, i.e. from France) French (Fr). It was predicted that generally duration and f0 must both be appropriate for a sentence to sound most rhythmically natural, but the precise ranges of acceptability for each cue may vary between language groups. The results provide evidence for these predictions, since stimuli with non-deviant duration and non-deviant f0 excursion were judged most rhythmically natural most often, and SG listeners were more tolerant of f0 deviation than duration deviation, whereas (S)Fr listeners were more tolerant of shorter than longer duration, and lower than higher f0 excursion.

The take-home message from these findings is that rhythm does not equal timing, since f0 (and probably other acoustic cues) are involved, and speakers hear rhythm differently depending on their native language. These conclusions have important implications for rhythm research in phonetics – it should not have a 'one-size-fits-all' approach. In other words, we should investigate: (i) not simply duration, but the interaction between various acoustic cues including f0, because these are likely to be interdependent in rhythm perception; (ii) perceived rhythm in speakers of various languages, to have a universal view of the phenomenon, unbiased by the relative importance of each cue in any particular language(s). A subsequent experiment took a first step in directly applying the perceptual experiment's results to production data, by developing a PVI (Pairwise Variability Index) which captures the multi-dimensionality and language-specificity of perceived rhythm in recorded speech.

What underlies the perception of rhythmic differences in speech?

Laurence White, Sven L. Mattys & Lukas Wiget
Department of Experimental Psychology, University of Bristol

Background. The perception of speech rhythm has been held to derive from the repetition of stressed and unstressed syllables, with cross-linguistic rhythmical differences arising, at least in part, from phonetic and phonotactic influences on the relative durations of stressed and unstressed syllables (Dauer, 1983). Spanish, for example, has relatively small differences in stressed and unstressed vowel duration, and few complex onsets and codas, whereas English stressed vowels are much longer than unstressed vowels, with stressed syllables also more likely to have complex onsets and codas. Rhythm metrics (Ramus, Nespor & Mehler, 1999; Low, Grabe & Nolan, 2000) are designed to quantify such differences, with two metrics particularly effective: VarcoV, the coefficient of variation of vowel duration, and %V, the proportion of utterance duration that is vocalic (White & Mattys, 2007).

The perceptual correlates of distinctions indicated by rhythm metrics have been investigated using utterances transformed into monotone sequences of *sasasa* syllables that preserve the durational values of the original vowels and consonants. Listening to *sasasa* speech, French speakers discriminated languages with distinct rhythm scores, such as English and Spanish, but not those with similar scores, such as English and Dutch, or Catalan and Spanish (Ramus, Dupoux & Mehler, 2003). However, *sasasa* speech contains potential cues beyond durational stress contrast: (1) Rate of syllable production: Spanish, for example, has simpler syllable structures than English, so more syllables will tend to be spoken per second even if rates of segment production are equivalent; (2) Intrinsic differences in stress distribution, e.g. fewer unstressed syllables between stresses in English than in Spanish.

Method and results. We used the *sasasa* resynthesis method, with all syllables converted to *sa*, and with consonant and vowel durations that of the original speech. F0 was a constant 230 Hz. We used an ABX categorisation task: on each trial, participants heard two randomly selected *sasasa* utterances, one from each language, and a third random sample – X – which could be from either language. The task was to decide which of the two languages the X sample was from. We tested 24 native English speakers for each of the four experiments.

Experiment 1: Replicating the study of Ramus et al. (2003), we compared Castilian Spanish (CS) and Standard Southern British English (SSBE), and found overall correct classification of 66%. Experiment 2: We again compared CS and SSBE, but with speech rate equalised between utterances. Classification accuracy dropped to 56%, showing that listeners exploit rate information, where available, to perform this task. Experiment 3: To eliminate differences in stress distribution, we compared two varieties of English, SSBE and Welsh Valleys English (WVE – which has rhythm scores intermediate between SSBE and Spanish). Rate was again normalised, and overall categorisation accuracy was comparable to Experiment 2. Experiment 4: We compared WVE with Orkney English, which has similar rhythmic properties to WVE. Categorisation accuracy was low, but still above chance.

Conclusions. Overall, the results clearly demonstrate that speech rate differences are important in the discrimination of rhythmically distinct language varieties. In Experiments 2-4, rate was normalised and categorisation accuracy was reduced. Furthermore, the importance of durational contrast between strong and weak syllables, as indexed by metrics such as VarcoV, is questioned here, as even rhythmically similar varieties of the same language can be marginally discriminated. Stress distribution does not appear a strong cue. Further analyses of individual test items suggest an important perceptual role for localised lengthening effects.

Selective prosodic marking in child-directed speech: a cross-linguistic study

Elinor Payne[1], Brechtje Post[2], Lluisa Astruc[3], Pilar Prieto[4, 5], & Maria del Mar Vanrell[d]

Phonetics Lab[1] and St Hilda's College, University of Oxford[1]; RCEAL[2] and Jesus College, University of Cambridge[2]; The Open University[3]; Universitat Pompeu Fabra[4]; ICREA[5]

A widely documented characteristic of child-directed speech (CDS) is the modification of certain phonetic parameters of prosody (e.g. higher and greater range of pitch; longer duration/slower speech rate; more prominent final lengthening; higher amplitude; and more even rhythm). A little studied question is the extent to which certain aspects of prosody may be prioritized or even suppressed. In other words, does CDS merely exaggerate the marking of prosodic structure to be found in adult-directed speech (ADS), or is it more selective?

As an initial foray into this issue, this paper examines the phenomenon of prosodic lengthening in English and Catalan. It asks the following research questions: i) is there evidence of greater prosodic lengthening in CDS?; ii) if so, are certain prosodic structures highlighted more than others; iii) and if so, are there cross-linguistic similarities in this, or is the selection more language-specific? We looked at the CDS and ADS of 3 Catalan-speaking and 3 English-speaking female adults, interacting with their 2-year old children and an adult interviewer. The material consisted of semi-structured dialogues elicited through short, animated clips shown on a computer.

Using Praat, the material was segmented into syllables, and the two most common syllable types, CV and CVC, were extracted and labelled according to level of prominence (lexically unstressed; lexically stressed but not accented; lexically stressed and accented; nuclear accented). Intonational phrase boundaries and word boundaries were also identified, and syllables labelled according to phrase position (initial, medial or final) and word position (initial, medial or final). A total of 1170 syllables were labelled, across language and speech style. Syllable durations, together with all other prosodic and word boundary information, were extracted using a Praat script.

A comparison of speech styles for each language suggest language-specific modifications in the parameters of lengthening for CDS. Specifically, unlike English, Catalan shows less durational variability as a function of syllable structure types in CDS than in ADS. Also, while in English syllables are longer in CDS roughly to the same degree across the word, in Catalan, they are longer in CDS specifically when word-final in a polysyllabic word, suggesting greater exaggeration of word-final lengthening in Catalan.

Lengthening patterns in ADS and CDS were then compared, for each language. In both languages, CDS presented a less *variegated*, more selective system of prosodic marking by duration than ADS. Specifically, English CDS appears to suppress phrase-initial lengthening and the distinction between stressed and unstressed syllables (away from the nuclear accented syllable), while prioritizing phrase-final lengthening. Catalan CDS appears to suppress lengthening of nuclear accents, while also prioritizing phrase-final lengthening. One result of this is that there is greater uniformity of syllable duration, which provides a probable explanation for the more even rhythm we observed for English and Catalan CDS. It remains to be investigated whether the suppression of durational signals reflect true systemic modifications, or if other phonetic cues, such as pitch movement, are present.

Is the phonetic encoding of grammatical boundaries based on form or function?

Mark J. Jones
University of Cambridge & University of York

Recent studies have identified grammatical structure as a source of phonetic variability. In Korean, palatalised nasals occur before /i/ within a morpheme but plain alveolar nasals occur before /i/ when a morpheme boundary intervenes (Cho 2001). In Lheidli consonants are longer across morpheme boundaries than across morpheme-internal syllable boundaries (Bird 2004). In English, varying acoustic effects are seen in prefixes (*mis*times) versus pseudo-prefixes (*mis*takes; Baker et al. 2007). Morpheme boundaries also affect Northern English glottalisation patterns (Jones 2007). The research presented here attempts to shed some light on the origin of the phonetic patterns encoding morphological structure. Some speakers lack phonetic indications of morphological boundaries altogether, so patterns could be emergent within an individual's grammar rather than acquired from other speakers. The patterns may therefore have little or no communicative intent, even if listeners can make use of them. The patterns themselves, when they do occur, may exhibit similarities across speakers if they represent automatic encoding of boundaries, perhaps caused by the need to retrieve and combine morphemes from the mental lexicon (cf. Marslen-Wilson 2001). The patterns involved may show similarities across speakers if they reflect a transparent (diachronic or synchronic or assumed) derivational relationship to other independent morphemes (e.g. English *not* > *n't*). On the other hand, detail could be more variable if it is functional in origin and speakers are individually motivated to provide information on morphological structure to interlocutors.

Results are reported here from an experiment into whether phonetic encoding of clitic boundaries operates along functional and derivational lines to differentiate e.g. 'Emma's (< has) rowed' from 'Emma's (< is) rowed'. Differences in vowel duration do occur, with longer vowels often present in the 'is' context. The existence of different vowel durations for different clitic contexts discounts a straightforwardly automatic origin of phonetic encoding. There is no evidence for a derivational origin based on 'has' or 'is' for differences in vowel quantity or quality. Most subjects pattern in the same way, suggesting that the details are not necessarily emergent. However, further work is needed to tease apart some additional questions. In particular, there is the possibility that the passive voice 'is' forms are less frequent and/or more ambiguous than the active 'has' forms and therefore require longer processing time which results in longer vowel durations.

F2-F3 amplitude relations and the perception of rhoticity in NURSE

Barry Heselwood & Leendert Plug,
University of Leeds

The marked downward shift of F3 observed on spectrograms during the production of rhotic approximants such as [ɹ] is often assumed to be important for the perception of these sounds. However, results from perception tests reported in Heselwood (2009) suggest that in the context of mid-central and back vowels F3, rather than facilitating perception of rhoticity, may inhibit it. These results have been explained by proposing that the auditory correlate of rhoticity is the presence in auditory space of a spectral prominence in the region of 9.0-11.5 Bark, and that this is best achieved if F2 inhabits the corresponding region of acoustic space (c.1080-1595Hz) without the presence of F3. When F3 is present in [ɹ], it is always within 3.5 Bark of F2 and therefore integrates with F2 to form a perceptual formant higher than F2 (Bladon, 1983). The frequency and/or bandwidth of this integrated formant seem to be less conducive to the triggering of perception of rhoticity than the lower frequency and/or narrower bandwidth of F2 on its own. It is suggested that the low amplitude of F3 typically found in [ɹ] (Stevens, 1998:542) is due to the vocal tract's attempt to filter it out so that it leaves F2 as undisturbed as possible.

Heselwood's explanation predicts that if there is no peak of prominence in the 9.0-11.5 Bark region, then there will be no perception of rhoticity. This paper reports two experiments designed to test this prediction. In experiment 1, using a rhotic token of *fir* with a stable formant pattern and average formant frequencies of F1=559Hz, F2-1444Hz, F3=1933Hz, a bandstop filter is used to vary the amplitude of F2 between -24dB and -0dB in eight 3dB steps. The steps are paired with the original unfiltered token and each pair presented in randomly varying order to phonetically-trained listeners who are asked if each member of the pair sounds rhotic, and if both do, which one sounds more rhotic. In experiment 2, using the same token of *fir*, a bandstop filter is used to vary the amplitude of F3 in the same way. The same procedure is followed as before. The bandwidth for both filters is equivalent to 4.88 Bark (800-1750Hz for suppressing F2, 1650-3470Hz for suppressing F3). The predicted result is that, firstly, listeners are least likely to perceive rhoticity when F2 amplitude is at its lowest, and most likely to perceive it when F3 amplitude is at its lowest; and secondly, that listeners will judge rhoticity to be greater the more that F2 amplitude exceeds F3 amplitude.

Centre-of-gravity calculations are made to obtain the value of the perceptual formant derived from the integration of F2 and F3 for each filtered token and its frequency and amplitude values noted for the perceptual threshold between rhoticity and non-rhoticity.

Multiple cues for the singleton-geminate contrast in Lebanese Arabic: the role of non-temporal characteristics

Jalal Al-Tamimi & Ghada Khattab
Speech and Language Sciences Section, Newcastle University

The phonetic and phonological aspects of gemination have been the subject of investigation in the literature and different definitions have been proposed. In Articulatory Phonology, for example, the difference between a geminate and non-geminate is said to depend on 'gestural stiffness' (Browman and Goldstein, 1990, 1992), whereby geminate consonants have less stiffness and thus longer durations. From a phonetic point of view, the geminate/singleton distinction is thought to be a difference in articulatory strength (Kohler, 1984) i.e. a fortis/lenis or tense/lax distinction (McKay, 1980). Fortis or tense consonants are produced with higher pulmonic strength and with more strength in their articulation, with longer duration and less voicing compared to lenis or lax consonants (Kodzasov, 1977; Jaeger, 1983, Ladefoged, and Maddieson 1996).

Although phonetic and phonological definitions may suggest qualitative and quantitative differences between geminates and singleton, most of the research on Arabic geminates has concentrated on the latter (e.g. Al-Tamimi, 2004; Ghalib, 1984; Ham, 2001; Hassan, 2003). In research on other languages, non-temporal characteristics have been proposed as secondary cues and are thought to enhance the perceptual distance between singletons and geminates (e.g. palatalised resonance for geminate sonorants (Local & Simpson, 1988), palatal contact for geminate stops (Payne, 2005), laminal contact for geminates as opposed to apical contact for singletons (Payne, 2006), lenited stops in singleton contexts (Ladd & Scobbie, 2004; Ridouane, 2007), and lower burst amplitude and occasional absence of bursts in singleton stops (Local & Simpson, 1999; Ridouane, 2007, forthcoming)).

This paper reports on the phonetic aspects of gemination in Lebanese Arabic (LA) and provides evidence for systematic qualitative differences between singleton and geminates in a wide range of spectral and other non-temporal cues not looked at in combination before. Twenty Lebanese males and females were recorded reading target word-lists containing medial singleton and geminate consonants (stops, nasals, fricative, laterals, rhotics, and approximants) preceded by long and short vowels. Acoustic and auditory analyses of medial consonants (C(C)) and of preceding (V1) and following (V2) vowels were made using Praat (Boersma and Weenink, 2009). Temporal measurements included V1, V2, and medial C(C) durations. Non-temporal measurements included formant frequencies at onset, mid-point and offset of V1, C(C) and V2; normalised intensity and RMS, *f*0 in V1, C(C) and V2; shape of the spectrum (Centre of Gravity and Peak) for stops and fricatives; duration of voiced and voiceless portions; and number of bursts (in the closure duration and in the VOT). Temporal results suggest a robust role for duration in distinguishing between short and long consonants and vowels in LA. Non-temporal results present a mixed picture and highlight the importance of looking at various cues for qualitative differences between singleton and geminate consonants in order to obtain a comprehensive picture of the phonetic implementation of articulatory strength. On the one hand, there were no differences between singletons and geminates with respect to Normalised Intensity, RMS or *f*0; on the other hand, significant differences emerged with respect to fewer voiced portions for geminate consonants compared to their singleton counterparts, and in the case of stops and fricatives, geminates exhibited significantly higher centre of gravity and peaks, and higher number of bursts and multiples bursts in the case of stops. These results suggest that obstruents may exhibit a starker singleton-geminate qualitative contrast than other consonants due to their more constricted manner of articulation, which is compatible with a stronger or more fortis type of articulation when duration is long, while showing lenition when duration is short.

Phonetic Cues of Psychologically Stressed Individuals in Forensic Contexts

Lisa Roberts
Department of Language and Linguistic Science, University of York & J. P. French Associates, York, UK

Speech and sound analysis has an increasing presence within criminal investigations. Given the widespread availability and use of mobile telephones, violent attacks are frequently audio-recorded by victims, witnesses and even the perpetrator(s) themselves.

The present study explores vocal responses of distress in real forensic situations where a victim has endured severe physical and emotional stress resulting from a violent attack. Previous studies investigating emotional speech have concerned speech technology applications (Erickson, 2003) and/or observed vocal cues of emotion in everyday speech (Scherer, 2003). Few attempt to characterise extreme emotion and fewer use authentic data.

The aims of the study are:

1. to identify vocal cues of authentic distress;
2. to investigate the limits of the individual's vocal performance.

Acoustic and auditory-phonetic analyses are conducted on the recordings - typically telephone calls to the emergency services involving an individual who has been - or is being - subject to a violent attack. Parameters under investigation have been chosen following observations from studies using recordings of psychologically stressed aviation personnel (William and Stevens, 1969; Kuroda et al., 1976), and include fundamental frequency (F0), tempo, intensity as well as vowel quality and intelligibility.

Findings show victims of violent attacks demonstrate extreme F0 increase and variability, fluctuations in tempo, and mixed results for intensity. Vowels also become more indeterminate as the attacks progress. A model is proposed charting changes to these vocal parameters relative to the timing of attack.

Quantity and Quantity Complementarity in Swedish and Arabic. A Phonetic Approach

Zeki Majeed Hassan.
Gothenburg University. & Al- Zaytoonah University.

Quantity and quantity complementarity play an important role in the phonetics and phonology of both Swedish and Arabic. Phonologically, vowel as well as consonant length is distinctive in Arabic and functions as part of the phonological system, though this is not so robust in Swedish and is still debated. Phonetically, on the other hand, the relationship between a vowel and a following consonant is one of inverse co-variation of duration in Swedish, but not so robust in Arabic and seem to function differently.

This study presents acoustic measurements of both vowel and consonant durations when long vowels precede geminates vs. singletons in Iraqi Arabic (IA) which have been overlooked by Hassan (2002 & 2003) to see if these durational differences operate differently from those when short vowels precede on the one hand, and on the other hand, from those in the same phonetic and phonological environments in Swedish. Contrary to the findings of Hassan (2002 & 2003), the durational difference between long vowels preceding geminates and those preceding singletons hardly showed any significant acoustic differences and perceptually go well under the DLS values. Also, geminates following long vowels are shown to be significantly shorter than geminates following short vowels. On the other hand, the present findings showed no significant inverse co-variation between V:C: whereas that between V:C showed very significant difference. However, the word overall duration showed no significant difference between words having V:C: and those having V:C, compatible with the findings of the two studies above. It is seen that quantity and quantity complementarity in IA and Swedish are language specific phenomena and operate differently in two different phonological systems. The study also poses serious questions on the predictability of length in Swedish phonology and whether gemination is lexical or post-lexical.

An exploration of the rhythm of Malay

W. Aslynn, G.J. Docherty, & E. Samoylova
Newcastle University, Speech and Language Sciences

In recent years there has been a surge of interest in speech rhythm. However we still lack a clear understanding of the nature of rhythm and rhythmic differences across languages. Various metrics have been proposed as means for measuring rhythm on the phonetic level and making typological comparisons between languages (Ramus et al, 1999; Grabe & Low, 2002; Dellwo, 2006) but the debate is ongoing on the extent to which these metrics capture the rhythmic basis of speech (Arvaniti, 2009; Fletcher, in press). Furthermore, cross linguistic studies of rhythm have covered a relatively small number of languages and research on previously unclassified languages is necessary to fully develop the typology of rhythm. This study examines the rhythmic features of Malay, for which, to date, relatively little work has been carried out on aspects rhythm and timing.

The material for the analysis comprised 10 sentences produced by 20 speakers of standard Malay (10 males and 10 females). The recordings were first analysed using rhythm metrics proposed by Ramus et. al (1999) and Grabe & Low (2002). These metrics (ΔC, %V, *r*PVI, *n*PVI) are based on durational measurements of vocalic and consonantal intervals. The results indicated that Malay clustered with other so-called syllable-timed languages like French and Spanish on the basis of all metrics. However, underlying the overall findings for these metrics there was a large degree of variability in values across speakers and sentences, with some speakers having values in the range typical of stressed-timed languages like English.

Further analysis has been carried out in light of Fletcher's (in press) argument that measurements based on duration do not wholly reflect speech rhythm as there are many other factors that can influence values of consonantal and vocalic intervals, and Arvaniti's (2009) suggestion that other features of speech should also be considered in description of rhythm to discover what contributes to listeners' perception of regularity. Spectrographic analysis of the Malay recordings brought to light two parameters that displayed consistency and regularity for all speakers and sentences: the duration of individual vowels and the duration of intervals between intensity minima.

This poster presents the results of these investigations and points to connections between the features which seem to be consistently regulated in the timing of Malay connected speech and aspects of Malay phonology. The results are discussed in light of current debate on the descriptions of rhythm.

Prosody in Hong Kong English: aspects of speech rhythm and intonation

Jane Setter

University of Reading

This paper reports research on prosodic features of speech in Hong Kong English (HKE), specifically, rhythm and intonation.

Setter (2006) adopted a pedagogically oriented, hierarchical methodology to examine HKE speech rhythm, in which weak, unstressed, stressed and nuclear syllables in a semi-scripted speech task were compared to British English data. It was found that rhythmic patterns in HKE differ significantly in comparison to British English, and it was hypothesised that this may lead to reduced intelligibility in international communication in some settings. In the current paper, the material from Setter (2006) is re-examined using the Pairwise Variability Index (PVI) (Low, Grabe & Nolan 2000), a measure which compares successive unit durations, in this case applied at the level of the syllable. Results show that a similar conclusion is reached using the PVI. In addition, new HKE data is presented to which the PVI is applied at the level of the syllable peak (vowel), and compared to findings from an existing study on British and Singapore English. It is found that the HKE data is more similar to the Singapore English data than the British English data.

Concerning intonation, this paper examines some features of tonicity and tone in HKE speech data, collected using an information gap task, in which the interlocutors are HKE speakers and a non-native speaker of English.

The results are presented in terms of patterns emerging in HKE as a World English.

Do rhythm measures separate languages or speakers?

Anastassia Loukina[1], Greg Kochanski[1], Elinor Keane[1], Burton Rosner[1] & Chilin Shih[2]
Oxford University Phonetics Laboratory[1], EALC/Linguistics, University of Illinois [2]

Ever since Pike and Abercrombie had suggested that all languages can be divided into stress-timed and syllable-timed, the so-called `rhythmic differences' between the languages have attracted substantial attention from phoneticians. Although experimental studies so far found no evidence for isochrony as such, various quantitative statistical indices have been proposed to capture the rhythmic properties of languages.

In this paper we compare 15 measures of durational variability based on an automatic segmentation of speech into vowel-like and consonant-like regions. Our corpus consisted of a total of 3059 short texts recorded from 62 speakers of Southern British English, Standard Greek, Standard Russian, Standard French and Taiwanese Mandarin. We used an automated algorithm to segment the data into vowel-like and consonant-like segments. This allowed us to apply identical segmentation criteria to all languages and to compute rhythm measures over a large corpus.

To compare intra-group variation in rhythm measures (RMs) to inter-group variation, we applied classifier techniques. We measured how often we can correctly predict the language, based on one or more RMs.

The performance of classifiers depended on the number of dimensions. While there was a significant difference in the performance of the classifiers based on single measures to classifiers based on three measures, there was only little improvement in the performance of classifiers based on more than three rhythm measures. This suggests that rhythm is at least a three-dimensional phenomenon and is best captured by a combination of more than two measures.

The most efficient classifier based on all 45 rhythm measures correctly identified the language of 61% of the data (chance=30%). This shows that although there are rhythmic differences between languages, substantial variation within languages makes it impossible to reliably separate languages based on the rhythm of a single paragraph.

At the same time, we have found that classifiers performed surprisingly well in identifying speakers of the same language. For example, for English classifiers based on three measures correctly identified the speaker of 48% of the data (chance=8%). Thus the differences between speakers of the same language appear to be more consistent than the differences between different languages. This finding raises interesting questions about the nature of individual variability in duration. It also shows that any future study requires a representative sample of speakers to avoid the danger of measuring differences between people rather than languages.

Factors Influencing Speech Timing in Non-fluent Aphasia

Anne Zimmer-Stahl, Rachael-Anne Knight, & Naomi Cocks
City University London

This paper sets out to discuss notions of non-fluency in aphasia and to clarify how they relate to phonetic measurements of speech timing. Previous research has linked non-fluent forms of aphasia to disturbances in speech timing. For example, Gandour et al. (1984, 1989 and 1994) found deficiencies in temporal control specifically in the coordination of simultaneous articulatory gestures. Furthermore, dysprosody has often been claimed to be a feature of Broca's aphasia, unfortunately making a somewhat impressionistic use of the term 'prosody'. So, when Baum & Boyscuk (1999) examined syllable duration with respect to utterance length, they were surprised to find no deviant results for non-fluent aphasics (except for a lack of phrase-final lengthening). However, such a result should not be surprising at all if we reflect on, firstly, what the different notions of speech timing stand for and, secondly, what wide array of communicative impairments is covered by the term 'non-fluent aphasia'. While, in practice, each patient seems to be different in how a lesion has affected his/her speech, in research we classify participants into syndrome groups resulting in a crude tendency to overgeneralize, thereby glossing over the issue of which factors have an impact on which parameters of speech timing.

This paper aims to disentangle factors that contribute to speech timing disturbances. Difficulties with *lexical retrieval* are contrasted with impaired *phonetic implementation* and are found to be captured by different measures of speech timing. It argues that we need to distinguish between (a) measures of *fluency* such as the CIU/minute (Correct Information Unit, Nicholas & Brookshire, 1993) or speaking rate/rate of articulation (Laver, 1994), and (b) measures of *articulatory timing* like VOT (voice onset timing). However, the syllable PVI (Pairwise Variability Index, Low, Grabe, & Nolan, 2000), as a relational measure, is not sensitive to fluctuations in temporal control for speech, but only shows deviant values in cases of genuinely *prosodic* impairment that inhibits the realization of speech rhythm.

The study draws on data from three mildly non-fluent aphasic speakers. Instead of comparing groups of aphasic speakers to a control group of non-aphasic speakers, a within-speaker analysis was undertaken. Aphasic speakers are known to show a great deal of variation in their speech behaviour from one day to another (or from the beginning of a testing session to the end). Texts (picture descriptions as well as free speech samples) from the same speaker yet differing in fluency (as manifested in the CIU/minute, a measure that can also be used to distinguish fluent from non-fluent types of aphasia) were analysed qualitatively for apparent production difficulties, and measures such as syllables/second, VOT, and syllable PVI were calculated. It can be shown that cognitive load has no effect on articulatory timing. Cognitive and word retrieval factors influence CIU/minute and speaking rate, but not rate of articulation or measures that reflect the ability to coordinate articulatory gestures such as VOT. The PVI seems not to be affected by within-speaker variability in fluency.

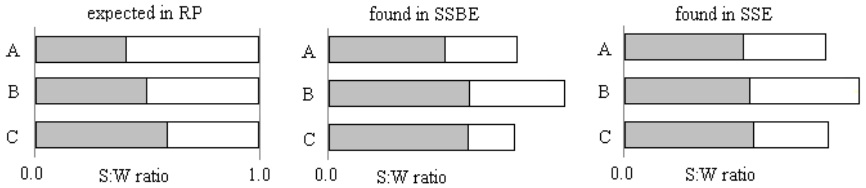
The results show that more clarity in the use of terminology is required when making claims about different aspects of speech timing in language and speech disorders. In particular, *prosody* as the phonological organisation of speech timing, *fluency* as the relationship between speaking rate and rate of articulation, and lower level *articulatory planning*, should be used as distinct units of terminology.

Foot timing in spontaneous speech: exploring rhythmic differences in two accents of English

Tamara Rathcke & Rachel Smith
University of Glasgow

Variation among accents of English in timing and rhythm has important potential implications for speech comprehension and conversational alignment among speakers of different accents. We investigate the production of such variation in two accents, Standard Southern British English (SSBE) and Standard Scottish English (SSE), focusing on 'micro-rhythm' or timing relationships within the foot. Specifically, we explore the empirical support for Abercrombie's (1964, 1979) distinction among trochaic foot types. Abercrombie (1964) proposes three foot types, with timing relationships as follows for RP: A, 'short-long' in *filler, city*; B, 'equal-equal' in *feeler, seedy*; C, 'long-short' in *feel a[live]*). The distinctions relate to the phonological weight of the words' first syllables, and the presence of a word boundary. Abercrombie (1979) further observes that B feet have a 'long-short' pattern in Yorkshire English, and a 'short-long' pattern in Scottish accents. Surprisingly, these observations have received little empirical attention.

We investigated spontaneous speech from middle-class speakers of two accents: SSBE spoken in Cambridge (http://www.phon.ox.ac.uk/old_IViE), and SSE spoken in Glasgow (Stuart-Smith, 1999). For each accent, 50 phrase-medial tokens of each of A, B, and C feet were selected, matched as far as possible for phonological composition, with (C)VCVC the majority pattern. We compared the durations of the strong and weak syllables, in both absolute and proportional (S:W ratio) terms. Medial consonants after lax vowels were treated as ambisyllabic, and half of their duration was assigned to each syllable.



The above Figure shows the patterns expected for RP (Abercrombie 1964), and our preliminary results. In SSBE, B feet were longer than A feet, but failed to show the expected greater S:W ratio. In SSE, B feet were again longer than A feet, with a smaller S:W ratio. Significant accent differences were found in B feet: the S:W ratio was smaller in SSE than SSBE, consistent with Abercrombie (1979). C feet showed large differences from A and B overall, but these were sensitive to phonological composition, suggesting that the C foot is not a homogeneous category. Considering only (C)VCVC tokens, we find a small difference between SSBE A and C feet, consistent with word-boundary-related lengthening (Turk and Shattuck-Hufnagel 2000). Long-vowelled C feet also had a smaller S:W ratio in SSE than SSBE, consistent with the B-foot pattern.

In summary, we found only partial support for the Abercrombian foot distinctions within either accent, but support for cross-accent differences, mainly reflecting the greater absolute and relative duration of unstressed syllables in SSE than SSBE. The data will guide design of a production experiment and a perception experiment testing the effect of timing disruption on lexical access.

Predicting prosody in poetry and prose

Greg Kochanski[1], Anastassia Loukina[1], Elinor Keane[1], Chilin Shih[2,] & Burton Rosner[1]
Oxford University Phonetics Lab[1], University of Illinois and the Beckman Institute[2]

Rhythm is expressed by recurring, hence predictable, beat patterns. Poetry in many languages is composed with attention to poetic meters while prose is not. Therefore, one way to investigate speech rhythm is to evaluate how prose reading differs from poetry reading via a quantitative method that measures predictability.

Our corpus consisted of 42 paragraphs read in each of Standard Modern Greek, Parisian French, Southern British English, Standard Russian, and Taiwanese Mandarin. Every reader also read 4 poems consisting of 8-12 lines. Where possible, we selected children's poetry with a regular metrical pattern that could easily be read with a strong rhythm.

We built a specialized speech recognition system, based on the HTK toolkit, that produced a sequence of **C** (consonantal), **V** (vowel-like) and **S** (silence/pause) segments. Once the segment boundaries were defined, five acoustic properties were computed for each segment: duration, loudness, frication, the location of the segment's loudness peak, and the rate of spectral change. We then computed 1085 linear regressions to predict these properties in terms of the preceding 1 to 7 segments.

For the regressions over the entire corpus, values of Pearson's r^2 varied widely: some regressions explain a negligible 2% of the total variance, and others up to 43%. Duration was the least predictable property: on average, r^2 was only 8%. This is noteworthy because all the published rhythm measures are based only on duration. The best predictable property was the rate of spectral change.

Overall, poetry was much more predictable than prose (r^2 values are roughly twice as large and our method allowed predicting up to 79% of variance). This is consistent with the intuition that poetry is more `rhythmical'. We also observed that poetry was more predictable across long ranges than prose. While in prose the mean difference between r^2 for the regressions based on 1 and 7 preceding segments was 6%, in poetry this difference was 25%. Given that all poetry in our corpus had regular metrical pattern, this confirms that the long-range effects we observe are likely to be related to such linguistic units as feet.

The predictability of a language depends on what is being predicted and the context of the target phones, so we anticipate that there will be at least several different ways to characterize the rhythm of each language. We propose that this approach could form a useful method for characterizing the statistical properties of spoken language, especially in reference to prosody and speech rhythm.

Cues to Vowels in English Plosives

Kaj Nyman
York University

Vocalic contrasts and their coarticulatory adjustments to consonants carry a great weight in the establishment of linguistic meaning (Hardcastle & Hewlett, 1998 and Coleman, 1998). The aim of the present study is to show to what extent listeners are capable of recognising English vowels from plosives. This question has important implications for the study of speech perception, phonology and coarticulatory theory.

A gating task experiment involving a forced choice method was used to assess how reliably listeners can recognise English monophthongs from voiceless plosives. 60 monosyllabic word stimuli embedded in a semantically neutral carrier phrase were selected. These were gated at the nearest zero crossings 10, 20, 30 and 40ms subsequent to plosive release. Four participants (from a set of 18 recordings) were selected as speakers for the perception experiment. All available participants were then asked to take part, while matching them according to VOT. Additional participants were also recruited to allow having a larger sample. The stimuli were played to participants in a random order, so that one out of four available answers had to be given for each stimulus. Although it was not possible to control for VOT with the second set of participants, the results suggest that the perceptual influence of VOT tends to be small (since participants' VOTs tend not to differ significantly).

The results show that all segmental constituents in monosyllabic English words have a significant bearing on vowels recognition. Velar and especially bilabial onsets lead to more correct recognitions than alveolar onsets, which require more precise tongue movement. Vowel quality (height in particular) also has clear implications on how vowels are perceived, so that high vowels are recognised much more reliably than low ones. The coda consonant also influences vowel perception, so that nasal codas give rise to a smaller proportion of correct responses than bilabial, velar and alveolar ones (respectively). The results show a good temporal progression in all of these cases, so that more acoustic information on a vowel leads to a larger number of correct responses.

The results show some of the limitations regarding our knowledge of speech (especially regarding coda consonants – also cf. Hardcastle & Hewlett, 1999). Not only do the results show that the phonetic interaction between plosion and vowel quality is essential for the proper perception of vowels (e.g. Liberman et al, 1967 and Tekieli & Cullinan, 1979), but that each segmental constituent in English CV(C) words contains a notable amount of phonetic information on the vowel. The phonological structure of a word thus significantly influences perception.

The results reaffirm key aspects of Articulatory and non-segmental (e.g. declarative) Phonology (e.g. Browman and Goldstein, 1986 and Coleman, 1998), since there is abundant evidence on feature-spreading, the perceptual influence of fine phonetic detail and coarticulation between segments.

An acoustic analysis of vowel sequences in Japanese

Isao Hara & Gerry Docherty
Newcastle University

Japanese is conventionally analysed as having five distinctive monophthong vowels and no diphthongs. However, all possible combinations of monophthongs into VV sequences are allowed and occur frequently both word-internally and across a word-boundary. Phonologically, two successive vowels such as [ai] are analysed as two syllables or moras (Inozuka *et al* 2003). However, there has been relatively little analysis of the phonetic characteristics of such sequences and how they might vary across different contexts. The few studies that have been carried out, however, suggest that the phonetic realisation of these sequences may be more complex than the phonological analysis suggests. Some studies, such as Saito (1997), note that Japanese vowel sequences can be diphthongs in fluent and fast speech. On the other hand, Gore (2006) measured the acoustic properties of one particular vowel sequence [ai] in three different morphological conditions, isolated production, morpheme-internal and across a morpheme boundary, concluding that there is little evidence for the [ai] sequence to be a diphthong in any of the conditions. Overall, the phonetic studies which have been carried out to date on Japanese VV sequences suggest that there is no consensus re: the extent to which it is appropriate to refer to these as diphthongs (Hattori 1967, Saito 1997, Kubozono 2001).

The aim of the present research project is to investigate the phonetic correlates of Japanese VV sequences in greater detail than has been achieved in previous studies. A wide range of VV sequences have been produced in a number of different environments and with different accent patterns by 6 male and 10 female speakers of Tokyo Japanese. Measurements include the durations of steady V1 and V2 intervals as well as the inter-vowel transition. Also comparisons have been made of the formant frequencies of vowels in VV contexts as opposed to when they occur as singletons.

This poster presents the results of a subset of the conditions investigated. They are discussed in light of whether the acoustic properties of vowels in VV sequences are significantly different from those of monophthongs, and whether the accent pattern and different phonological contexts have a role to play in respect of the acoustic properties investigated.

Improved representation of variance in measures of vowel merger

Lauren Hall-Lew
University of Oxford

Previous measures of vowel merger, such as the Euclidean distance between averages, have only been able to capture some of the variability between two given vowel clusters. Reliance on averages obscures the amount of variability within a given vowel class, while other techniques, such as calculating distance between minimal pairs, rely on few tokens per speaker. Both cases reduce statistical power and reliability. Hay, Nolan and Drager (2006) introduced an alternative approach that accounts for the variability between two vowel clusters and only requires formant values as input, rather than averages. The measure, known as a Pillai score (or the Pillai-Bartlett statistic; Baayen 2008:158), is the output of a Multivariate Analysis of Variance (MANOVA). The Pillai statistic is a numerical representation of the proportion of variance in one cluster that can be predicted by the variance of the other cluster. A higher value indicates a lower degree of overlap between the two clusters in F1/F2 space, so, “the lower the Pillai score, the more advanced the merger” (Hay et al. 2006:467). Since the value is derived from a MANOVA, the Pillai score can easily account for known internal factors influencing the production of merger, such as phonological environment, thereby reducing the need to obtain minimal pair lists.

This talk argues for using Pillais as measures of merger by comparing the analysis of LOT/THOUGHT merger in California English with that of NEAR/SQUARE merger in New Zealand English (Hay, Nolan, & Drager 2006), considering the consequences of using Pillais with respect to *mergers-by-approximation*, *mergers-by-transfer*, and *mergers-by-expansion* (Trudgill & Foxcroft 1978; Herold 1990). The talk also presents a new application of Pillai scores as measures of back vowel fronting in California English, since in representing the difference between any two clusters, the statistic can represent any measure of variable distance in vocalic space.

Reliability of formant measurements from lossy compressed audio

James Bulgin[1], Paul De Decker[1] & Jennifer Nycz[2]
Memorial University[1], University of York [2]/ New York University[2]

Lossy audio compression algorithms, such as those used in mp3s and VoIP services like Skype, achieve their high levels of compression through destructively modifying the original waveform. Although lightly compressed audio is indistinguishable from uncompressed audio to the human ear, it is unclear how strong an effect compression might have upon the accuracy of acoustic measurements made for the purpose of phonetic study. This study examined formant measurements from a series of sociolinguistic recordings of both male and female speakers, and their reliability when these recordings were compressed.

If acoustic analysis of compressed audio is sufficiently reliable, it could provide practical benefits to researchers. For instance, some collections of linguistically interesting recordings may not be available in a lossless format. In addition, the much smaller file size of compressed audio (potentially 10-30 times smaller than uncompressed audio) could simplify the management of large corpora, and make it more feasible to share them with other researchers, especially over the internet. Finally, VoIP services could offer a potentially useful tool for gathering linguistic recordings from remote locations.

In this study, recordings originally encoded as 24-bit, 44Hz wav files were re-encoded as mp3s, at three levels of compression (64, 128, and 320 kbps (CBR)) using the encoder built into Sound Forge. Additionally, the originals were transmitted and re-recorded over Skype, to test the compression algorithms used internally by this program. F0 though F4 were measured using Praat (Boersma and Weenink 2009) at the temporal midpoint of approximately 100 vowel utterances for two speakers, and these measurements were repeated at the same timestamps for each version of the recording. The results for each compressed version were then compared with the original measurements.

Results suggest that even high levels of mp3 compression have a minimal effect on the accuracy of formant measurements (a result similar to Van Son 2005). For the speakers examined, F1 and F2 for each vowel type differed from the original recording by an average of 3Hz and 9Hz, respectively, on average. For many linguistic purposes, this is an acceptable margin-of-error. Recordings transmitted over Skype differed from their originals to a significantly greater degree, and it does not appear at this time to be a suitable tool for gathering recordings where accurate acoustic analysis is required.

Static and Dynamic Cues in Vowel Production and Perception: Cross-Linguistic Investigation of Moroccan Arabic, Jordanian Arabic and French

Jalal Al-Tamimi
Laboratoire Dynamique du Langage, Université Lyon 2, France, and Speech and Language Sciences Section, Newcastle University

This paper reports on the role static and dynamic cues obtained from production data and used by listeners of Jordanian Arabic, Moroccan Arabic and French. The ongoing debate regarding the role of static and dynamic cues in vowel production and perception suggests that the use of dynamic cues, such as VISC measurements (Nearey and Assmann, 1986; Nearey, 1989; Hillenbrand et al., 1995; Hillenbrand et al., 2001); the Silent Center Syllables (Strange et al. 1976; Strange, 1989, 1999, etc), or formant trajectories (van Son and Pols, 1992; Pols and van Son, 1993; McDougall, 2006), enable the listeners to better identify vowels obtained from Isolated or from syllable like production, as proposed by the proponents of the Dynamic Specification Theory (Strange, 1989, 1999; Carré, 2004, 2006; Lindblom et al. 2006; *inter alia*). The explanation provided is that the acoustic properties included in formant trajectories provide important clues to vowel identity.

To test this theory, three different languages are compared: Moroccan Arabic (with a 5-vowel system /iː ə aː uː/, (Hamdi, 1991); Jordanian Arabic (with an 8-vowel system /iː i eː a aː oː u uː/, Bani-Yasin, & Owens, 1987) and French (with an 11-vowel system /i e ɛ a ɑ ɔ o u y ø œ/), (MA, JA, FR, henceforth). Phonetic-phonological, morphological and lexical (between MA and JA) differences between these three languages enables us to test the (dis)similarities in the use of static and dynamic cues in both production and perception.

In Production, 10 male speakers per language (aged 20 to 30) were recorded reading a list of vowels in C_1VC_2, C_1VC_2V, and C_1VC_2VC, where C_1 and C_2 were either /b/, /d/ or /k/, and V, each vowel. The speakers were asked to produce these items with normal rate and unmarked style. Formant frequencies at the temporal mid-point were determined to represent static cues, while a smoothing of formant transition from the onset through to the temporal mid-point (via linear, quadratic and cubic regression analyses) was used for dynamic cues. In perception, a prototype identification task based on a Method of Adjustment (Johnson et al., 1993) was used, whereby the entire F1~F2 acoustic vowel space is a continuous synthesised space and enables listeners to obtain all the possible synthesised stimuli from the values of formant frequencies. Static cues were proposed as sustained vowel targets without any inherent movements, while dynamic cues were a combination of a sustained vowel target preceded by a transition (or a CV syllable, where C was /b d or k/). 10 male listeners per language (aged 20 to 30) were asked to identify the prototypes of their system, in V and CV stimuli embedded in words.

Results from the production task showed that the use of linear regression analysis on each formant of the vowels in the three languages enabled a better discrimination of the vowels, within and between the languages (using both a factorial MANOVA and a Discriminant Analysis), and suggested that dynamic cues enable to better describe the differences between and within the languages. In perception, the use of dynamic cues (or CV syllables) enabled listeners of the three languages, but more particularly JA, and to some extent, MA, to perform better in the identification task. In FR, the use of either static or dynamic cues enabled the listeners to identify correctly the prototypes of their system.

These results suggest that the use of dynamic cues in association with static ones enables the listeners to better identify the vowels of their system and show that the cues used in perception to identify vowel depend on the Dynamic Specification of vowels and may be both static and dynamic (e.g. Strange, 1999; Carré, 2009).

The contribution of accent distribution to accentedness: causes and implications

Sam Hellmuth
University of York

Although in English, in certain contexts, a speaker might realise an utterance with a pitch accent on every word, in most contexts, in naturally occurring speech, speakers realise pitch accents every few words or so at most. The distribution of accent in English is known to interact with information structure (Gussenhoven 1983, Selkirk 1984) and can be formalised in terms of phrase level accent distribution (Selkirk 2000). There is evidence that Egyptian Arabic (EA) displays a very different accent distribution pattern, with a default pitch accent observed on every content word, in a range of speech styles (Rifaat 1991, Hellmuth 2006). A similar rich accent distribution pattern has been described for other languages including Spanish, Greek (Jun 2005), Hindi (Harnsberger 1996, Patil et al. 2008) and Tamil (Keane 2006), and, crucially, in varieties of English with a 'frequent accent' substrate, such as Tamil English (Wiltshire & Harnsberger 2006). In the light of this last fact, we hypothesise that rich accent distribution will transfer into the L2 English of L1 EA speakers, and present here the results of two studies exploring this prediction and its implications.

Firstly, we present the results of a production study which documents the accent distribution patterns in the L2 English speech of two female speakers of EA, who recorded the IViE corpus stimuli (http://www.phon.ox.ac.uk/IViE), with comparison to the accent distribution pattern observed in the speech of two female L1 speakers of Southern British English from the IViE corpus, in parallel utterances. We observe a higher incidence of accents in the EA speakers' L2 English, across a range of utterance types and styles.

Secondly, in the light of anecdotal evidence which suggests that speakers of 'infrequent accent' languages like English commonly perceive speakers of 'frequent accent ' languages such as EA to be speaking in an angry or aggressive manner, we present the results of a pilot perception study which seeks to determine how frequent accent distribution is interpreted by English listeners. A possible explanation of this anecdotal evidence is that the different accent distribution patterns in different languages map to a different function: in rich accent distribution languages, the pitch movement serves only to mark word-level prominence (Jun 2005, Hellmuth 2007), so that the mere occurrence of an accent provides no contribution to meaning; this contrasts strongly with the function of accent distribution which, as noted above, serves to mark argument/information structure. According to the Effort Code (Gussenhoven 2004:79), increased incidence of pitch movements will be interpreted paralinguistically as emphasis or insistence. If a speaker produces 'unnecessary' extra accents in their L2 English, due to transfer of a linguistic accent distribution pattern from their L1, there is a risk that this will be interpreted paralinguistically instead of linguistically. We manipulated the incidence of pitch movements in two sample utterances, in EA and English, and they were rated separately for degree of emphasis and degree of insistence by 4 English listeners and 4 EA listeners (with advanced L2 English). The results indicate that both L1 English listeners and advanced L2 English EA listeners interpret utterances containing a greater incidence of pitch movements as more insistent.

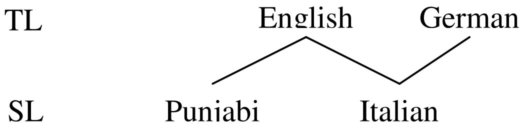
A ‘basic variety’ in learner intonation and its development over time

Ineke Mennen[1], Fredrik Karlsson[2], & Aoju Chen[3]

ESRC Centre for Research on Bilingualism, Bangor University[1], Department of Language Studies, Umeå University[2], Max Planck Institute for Psycholinguistics, Nijmegen[3]

Contrary to most studies on L2 intonation which take a ‘target deviation’ approach and focus on how one or two aspects of L2 intonation diverge from the target language, this study took a ‘learner variety’ approach (Klein & Perdue, 1997). In a ‘learner variety’ approach L2 learner intonation varieties are regarded as systems in their own right and are analysed along four dimensions of intonation (Ladd, 1996): (i) the inventory of structural elements (pitch accents and boundary tones); (ii) the way these elements are realised; (iii) their distribution; and (iv) their use to signal intonational function (e.g. to mark interrogativity). This approach ensures that each learner intonation variety is analysed in its entirety, rather than providing just a partial description of certain aspects of intonation. This allows insights into the acquisitional process of intonation. Furthermore, the longitudinal set-up of this study allows us insights into whether and how an intonational system evolves over time.

We selected statements and questions from a subset of the European Sciences Foundation L2 Database, covering the longitudinal data of two groups of speakers from a structurally different source languages (Punjabi and Italian) learning the same target language (English), and two groups with the same source language (Italian) learning a different target language (English and German), as exemplified in the figure below. This design enabled us to determine whether learner varieties develop over time and whether there are source-language-independent patterns in acquiring the intonational system of a L2.



Results show surprising similarities across speakers in most dimensions of intonation, in particular in the inventory and functional use of intonation, regardless of the source or target language. Source language influences were confined to the distribution and phonetic realisation of some pitch accents and/or boundary tones. Longitudinal development was also similar for the different learner groups, and was also most apparent at the level of phonetic realisation and distribution. We propose that learners start out with a set of basic elements (‘basic variety’) that build a simple but fairly efficient intonation system. We will describe the characteristics of this basic variety and how it evolves longitudinally, and will briefly discuss how our results help inform knowledge of the acquisitional process and phonetic/phonological learning.

Combining prosodic cues to signal topic structure

Margaret Zellers & Brechtje Post
Research Centre for English & Applied Linguistics, University of Cambridge

A variety of prosodic cues have been identified as signals to the organization of spoken discourse around topics, or topic structure. These cues have included the height and possibly the temporal alignment of fundamental frequency (F0) peaks (Wichmann 2000), the length of pauses between utterances (Swerts & Geluykens 1993), the amount of variation in overall F0 range in utterances (Nakajima & Allen 1993). Previous work by the authors also identified the range of F0 falls, and variation in speech rate at the beginnings of utterances, as showing variation in parallel with the topic structure (Zellers & Post, to appear). In addition to those phonetic variations which seem to be directly linked with topic structure, there also appear to be changes in the distribution of other features which are connected with topic structure but not directly dependent on it. In previous research, we found that F0 peak alignment differences in different topic structure categories may actually signal two different pitch accent categories, which have different distributions on the basis of topic structure (Zellers et al. 2009). We also found that the distribution of aperiodicity at the ends of utterances seems to bear at least some relationship to the topic structure of the discourse, although only for certain category contrasts (Zellers & Post, submitted). The distribution of these features, although not able to directly signal contrasts in topic structure category, is still likely to aid listeners in identifying topic structure.

Emerging from this assortment of prosodic signals is a picture of a complex system in which many cues appear simultaneously to create variations in prominence, perhaps through an additive effect, or by varying more or less from some neutral baseline (as suggested by Xu 2005 for speech production in general). In this paper, we will argue that these changes in phonetic prominence lend themselves to the signalling of topic structure in spoken discourse. We will examine some of the ways in which the prosodic cues mentioned above can interact to create prominence variations around different levels of topic structure, focusing particularly on the interplay between intonational or F0 cues and other aspects of prosody.

Prominence, stress and timing in Evenki

E. Samoylova
Newcastle University, Speech and Language Sciences

Evenki is a little studied Tungus language that is spoken predominantly in the north-east and east of Russia (Boldyrev, 2000), as well as by minority communities in Mongolia and China (Bulatova & Grenoble, 1998). Although some work was done on Evenki syntax and morphology (Kolesnokova, 1966; Kilby, 1980; Nedyalkov, 1997), phonetic properties of this language have scarcely been studied at all. Limited information about general phonemic inventory, and allophonic distributions and permissible syllable structures of the language is available (Nedyalkov, 1997; Bulatova & Grenoble, 1998; Boldyrev, 2000); however very little work has been carried out on aspects of prosody. Furthermore, to the best of the author's knowledge Evenki phonetic system in general and prosodic features in particular have never been studied instrumentally.

A preliminary pilot study was conducted that aimed at providing a description of prosodic features associated with prominence and timing on the basis of instrumental analysis.

The data for the analysis comprised the recordings of 3 native speakers (2 males and 1 female) reading the list of individual words as well as 30 sentences containing those words. Mean f0, mean intensity and syllable duration as well as the intensity and duration of individual segments in the syllable were obtained.

This poster will present the results of this study. In particular, it will address the question of whether there is lexical stress in Evenki.

Marking of information structure in semi-spontaneous speech: Southern British English and Seoul Korean

Hae-Sung Jeon[1] & Margaret Zellers[2]

Department of Linguistics[1], Research Centre for English and Applied Linguistics[2], University of Cambridge

Language users in general seem to make important information in speech more acoustically salient compared to less important elements, but the way this is achieved in the acoustic signal varies from language to language. This variation gives rise to differences in the phonological description of their prosody (e.g. Frota 2002). In this paper, we investigate phonetic variation with relation to the production of focused constituents in Standard Southern British English (SSBE) and Seoul Korean (SK), languages that vary greatly in their phonological descriptions. By using the same methodology across languages we are able to make direct comparisons between the two languages.

Speech data were collected from native speakers of SSBE and SK. We used two game tasks played by a pair of speakers, replicating studies by Swerts et al. (2002) and Hellmuth (2005). In the first task, speakers were given a set of cards with 8 different combinations of colours and shapes. Their task was to make identical lists of the 8 cards without seeing each other's cards. In this process, they produced noun phrases with a fixed construction, e.g., "green circle," and each of the two constituents was designed to be new, contrastive, or given information at different points in the game. The second game was a modified version of the 'murder mystery' game used by Hellmuth (2005). In this game, which was also played with sets of cards, speakers were asked to collaborate to identify the right combinations of murderer, location and weapon by using sentences, e.g., 'it was the *person* in *place* with the *weapon'*.

We will address whether the different information status can systematically explain acoustic variations in both languages and to what extent the variations appear similarly across the two languages, which have been given very different phonological descriptions (cf. Pierrehumbert 1980 for English, Jun 1998 for Korean). We will present the results of the preliminary analyses of word duration, distribution of fundamental frequency (F_0) peaks, and F_0 range of prosodic phrases.

Suprasegmental differences between the voices of preadolescent girls and boys

Sharon Phillips, Ghada Khattab & Gerry Docherty
Speech and Language Sciences Section, Newcastle University

Previous research has revealed gender judgment accuracy for preadolescent voices to be above chance (Perry et al. 2001). Attempts to determine what acoustic and perceptual factors differ between the voices of girls and boys, however, have yielded inconclusive results, especially with regards to suprasegmental cues (Weinberg and Bennett 1971; Sungbok et al 1999). Our aim is to investigate the potential differences between the speech of preadolescent boys and girls on a range of non-segmental cues and to explore how these differences change as a function of age. Thirty typically-developing, British-English speaking girls and boys (ages 6, 8 and 10) were recorded producing single words and spontaneous speech. Their voices were analysed for gender differences on measures of f_0 (mean, median, maximum, minimum, *SD,* range), speech rate and intensity.

The various acoustic measures of f_0 were found to be significantly different between the girls and boys in all age groups. These are surprising and interesting results, especially for the 6-year-old group. In previous research, $f_0$has not shown to act as a cue for gender distinction in preadolescents, and researchers have attributed that to the absence of significant physiological differences in the vocal tract until age 12 or after (Gunzburger et al 1987; Perry et al. 2001; Sungbok et al 1999). Gunzburger et al (1987), for instance, compared the length and circumference of the throat for girls and boys aged 7 to 8 and found no significant sex difference. There may, however, be other relevant physical differences between the two sexes, even at this young age. This has been shown in research on vowel formants in children, where differences between boys and girls can be attributed to neck circumference in preadolescent girls and boys (Bennett 1981). Moreover, Kahane (1981) found that male vocal folds increase in size at more than twice the rate of the female vocal folds from preadolescence to puberty. This may explain why results for mean, median, maximum and minimum f_0 for the girls' fundamental frequency did not vary much across age groups, whereas the boys' fundamental frequency significantly decreased as age increased from 6 to 10. Ferrand & Bloom (1996) and Sungbok et al. (1999) have also reported more gradual f_0 changes for girls compared to boys.

Another tentative explanation for the results in this study is that children may be subconsciously altering suprasegmental features of their voice to conform to a gender stereotype. In Bennett's (1981) study, for instance, the formant frequency differences that were found between boys and girls were still prominent even after body size was taken into account. The differences were attributed to learned behaviour in boys and girls, such as extent of mouth openings and lip rounding. Acoustic aspects which are different between the sexes (e.g. vocal pitch/quality) may therefore be incorporated into children's gender schema which could consequently result in them adapting their voices to sound more feminine/masculine. Further evidence from the current study can be found in the results for intonation range, whereby girls of all age groups were found to have a wider pitch span and more varied *SD* than boys, especially in spontaneous speech. This supports previous literature that has found that higher intonation variability is perceived as a female attribute (Bennett & Weinberg 1979; Gunzburger et al.1987; Wolfe et al. 1990). Speech rate was found to significantly increase after age six, with the 8- and 10-year-olds having considerably faster speech rates, but there were no significant gender differences in speech rate at any age point. Finally, in terms of intensity, the current study girls were found to have acoustically quieter voices than boys, though this was only significant at age 8.

An interactional approach to prosody: variation and constraints in English news receipts

Marianna Kaimaki
Department of Language & Linguistic Science, University of York

In this paper I argue that prosodic variation:

a) may be determined by interactional organisation with rises and falls being in free variation at particular places in interactional structure;
b) may be conditioned by turn-design.

Interactional and phonetic analysis of news-telling sequences in English telephone calls suggests that the choice of falling or rising pitch contour is not consequential for the design or subsequent development of the talk. News-receipts with falling and rising pitch contours may have the same uptake and be treated in the same way by co-participants. I argue that in such sequences, these contours can be considered to be in free variation.

I also show that turn-design constrains pitch characteristics such that news-receipts formatted with *oh really* or *oh good* are associated with both rising and falling pitch contours while news receipts done with *oh+adj* (where *adj* ≠ good) are only associated with falling pitch contours.

The effect of r-resonance information on native and non-native speakers of English

Antje Heinrich & Sarah Hawkins
Department of Linguistics, University of Cambridge

An r-segment in an English word typically produces formant frequency changes for up to about 600 ms before the r-segment itself (Heid & Hawkins, 2000). These changes are known as r-resonances. Previous listening experiments have shown that long-domain resonances are perceptually salient for native speakers of English in synthetic speech (Hawkins & Slater, 1994), single carrier phrases (West, 1999) and unrestricted read sentences (Heinrich & Hawkins, submitted). Whether they are salient for non-native speakers of English is unclear.

The current study compared 27 native speakers of German with 24 native speakers of English in their use of r-resonances to aid intelligibility of read sentences heard in background noise. Test material comprised 52 pairs of English read sentences that only differed in the presence of an /r/ or an /l/ in a minimal-pair target word (e.g. *mirror, miller*); neither /r/ nor /l/ occurred elsewhere in the sentence (the base). Target words were cross-spliced into a different utterance of the same sentence base (match) and into a base that had originally contained the other member of the minimal pair target (mismatch). All sentences were heard in 12-talker background noise. Each listener heard only one member of each pair. Percentage of words correct for spliced target words and selected unspliced words in the preceding base was measured.

The native English listeners showed a strong facilitating influence of r-resonance in the sonorant segment immediately preceding the critical /r/ segment, and a weak facilitating influence from the long-domain resonances in the preceding base. The native German listeners used only the long-domain r-resonances, and not the stronger short-domain ones, and only when the listener was relatively inexperienced with English. German listeners with several years' experience of UK English ignored r-resonances in the speech signal completely, possibly in favour of lexis and syntax. Relatively inexperienced German listeners' failure to use the strong resonance information in the immediately adjacent sonorant is presumably because the English /r/ articulation is quite different from that of German uvular /r/ so did not meet expectations for an /r/. Their use of the weaker long-domain resonances has yet to be elucidated. As a group, the German listeners were more disadvantaged by the splicing itself than were the English listeners, indicating the fragility of accurate speech perception in a non-native language. These results suggest complex interactions between acoustic-phonetic and 'higher level' linguistic knowledge in the process of proficient L2 acquisition.

Influence of r-resonances on speech intelligibility of young and old listeners

Antje Heinrich & Sarah Hawkins
Department of Linguistics, University of Cambridge

Declining hearing sensitivity in high frequencies and impaired speech understanding in noise are hallmarks of auditory aging. But not all abilities decline with age: knowledge about language tends to improve. Consequently, it is possible that older listeners compensate for a decline in hearing by increased reliance on linguistic knowledge. The current study is part of a programme of research into age-related interactions between the auditory signal and top-down phonetic and semantic knowledge. The phonetic property investigated in the current study is ‘r-resonances’. Resonances are interesting to investigate because they provide information about the phonetic content of an upcoming speech segment if the listener knows how to interpret the acoustic cues. In young listeners, r-resonances have been shown to be perceptually salient in adverse listening conditions when added to synthetic speech (Hawkins & Slater, 1994), in natural read speech in a single carrier phrase (West, 1999) and in unrestricted sentences (Heinrich & Hawkins, submitted). Whether older listeners with age-related hearing loss can make use of r-resonances is unclear. However, r-resonances should be useful to older listeners because their acoustic characteristics lend themselves to listeners with presbycusis: r-resonances are fairly long-lasting (up to about 600 ms before an /r/ segment) and of relatively high amplitude below 4 kHz.

Test material comprised 52 pairs of English read sentences that only differed in the presence of an /r/ or an /l/ in a minimal-pair target word (e.g. *mirror, miller*); neither /r/ nor /l/ occurred elsewhere in the sentence (the base). Target words were cross-spliced into a different utterance of the same sentence base (match) and into a base that had originally contained the other member of the minimal pair target (mismatch). All sentences were heard in 12-talker background noise. Each listener heard only one member of each pair. Percentage of words correct for spliced target words and selected unspliced words in the preceding base was measured. Forty-eight old listeners (age range 60-80 years), with varying degrees of hearing loss, typed what they heard, prompted by the first few words of the sentence. The results were compared with those for 24 normal-hearing listeners aged 18-25.

Preliminary analyses show that young listeners and older listeners with relatively good hearing use r-resonance information and not word frequency. However, the effectiveness of r-resonance information for older listeners depended on their degree of hearing loss. Only those with better hearing used the information in similar ways to young normal-hearing listeners. Older listeners with a greater degree of hearing loss used lexical knowledge in addition to resonance information: for such listeners, higher frequency words were more likely to be correctly identified than low-frequency words. In a next step, we will low-pass filter the speech stimuli such that frequency information below the hearing threshold of the hearing impaired group is inaudible. These filtered sentences will be presented in background babble to young normal-hearing listeners to see if they adjust their listening strategy to rely more word frequency information, like the older hearing-impaired listeners. Data for filtered sentences will be presented along with that already collected for the unfiltered stimuli.

A phonetic study of the rhotics in Malayalam

Reenu Punnoose
University of Newcastle-upon-Tyne

Malayalam, a Dravidian language spoken in Kerala, South-West India has a five member liquid inventory: two rhotics, two laterals and a fifth liquid that has been variously referred to as an approximant, fricativised lateral, ‘r-sound’, continuant etc. The phonetic and phonological nature of the two rhotics is inconclusive in the limited literature that exists on the language. Some categorize the two as two ‘trills’ (Ladefoged and Maddieson, 1996; Srikumar and Reddy, 1988), others as one tap and one trill (Kumari, 1972; Yamuna, 1986) etc. Furthermore, a recent study of the Brahmin dialect of Malayalam’s parent language Tamil (McDonough & Johnson, 1997) revealed that the fifth liquid common to both languages was a third rhotic in the latter’s inventory. The rhotics in particular are interesting because they are lexically contrastive. For example, /kari/ (soot)-/ kari/(curry).

The author addresses two main questions: Firstly, what are the phonetic cues that speakers use in order to maintain the lexical contrast between the two uncontested rhotics? Secondly, could the fifth liquid be a potential third rhotic in Malayalam, and what are its auditory and acoustic characteristics? Eight male and eight female speakers were recorded producing words containing the five liquid consonants in a carrier phrase. Auditory and acoustic analyses were conducted on the data. Preliminary results suggest all tokens of one rhotic are produced by all speakers as taps whereas the productions of the other rhotic varied across speakers between a tap or a trill. Target trills irrespective of whether they were realized as such or not and their surrounding vowels were realized as more retracted and more open than target taps and their surrounding vowels, which were more fronted and more close. Apart from the difference between the consonants, the difference in the auditory quality of the surrounding vowels appears to be a strong distinguishing cue between tap-trill minimal pairs. These auditory characteristics are also reflected in their corresponding acoustic features: Taps and surrounding vowels seem to show a significantly higher F1 and lower F2 and trills and surrounding vowels seem to show a significantly lower F1 and higher F2. The fifth liquid tokens sound like a post-alveolar approximant, almost retroflex. However, vowels preceding it, front or back, sound fronted and raised, contrary to what is expected in a retroflex environment. Acoustically, it exhibits low F3 values and high F2 values and compared with the rest of the four liquids, it tends to have the closest proximity to lamino-alveolar tap and lamino-alveolar lateral approximant. Duration was found to be a significant cue with respect to the taps and trills but not their surrounding vowels. Target trills and trill realizations were found to be longer than target taps and tap realizations. These results suggest that speakers use duration of the consonants and systematic differences in the surrounding vowel quality (fronting vs. backing; close vs. open) and between the consonants (advanced vs. retracted; laminal vs. apical) to distinguish between tap-trill minimal pairs. The auditory and acoustic features of the fifth liquid suggest one of several possibilities: it could be a palatalized post-alveolar approximant or a *palatalized* ‘retroflex’ approximant or perhaps a *palatal* approximant. Another Dravidian language, Toda, is reported to have a palatalized retroflex trill in its inventory (Spajic et al.,1996). This raises interesting questions since recent studies (Hamann, 2003) have argued that palatalized ‘retroflex’ sounds cannot exist due to the opposing nature of the articulations involved In palatalization and retroflexion. Work is in progress for further detailed analyses.

Acoustic characteristics of /r/ produced by people with hearing impairments

Rebecca Coleman, Rachael-Anne Knight & Ros Herman
City University London

There is convincing evidence that normal speech perception is auditory and visual in nature (see Woodhouse, Hickson and Dodd, 2009, for a review). Pre-lingually Deaf people, however, must develop speech primarily by using the visibility of phonetic features, and their lip-reading abilities are enhanced compared to hearing controls (Auer and Bernstein, 2007).

As [ɹ] is often produced with some degree of lip rounding, it is hypothesised that Deaf speakers will focus on this visible articulatory characteristic, rather than the acoustic cues they cannot hear. As a result, it is likely that a Deaf person will produce a heavily labialised or labiodental /r/. The present study aims to investigate the acoustic characteristics of /r/ produced by Deaf Speakers.

The participants for the study were three deaf speakers and three hearing speakers matched on age, sex and regional accent. Each participant read a list of phrases containing /r/ and /w/. Formants were measured and the average frequencies were
then compared between the Deaf and hearing participants.

The Deaf speakers had, on average, higher third formants than the matched controls, at frequencies characteristic of a labiodental articulation. The matched controls had low third formants and small distances between F2 and F3, which is characteristic of an apical articulation. The results also showed that one of the Deaf participants made no significant acoustic difference between his production of /w/ and his production of /r/ (although it is possible that he makes a covert contrast which would only become apparent using other acoustic measures).

These findings suggest that Deaf speakers produce labiodental /r/s due to their reliance on visual cues. Implications for theories of audio-visual speech perception are discussed, as is the possibility that hearing speakers who use labiodental /r/ might also rely more heavily on visual cues during speech acquisition.

Non-segmental correlates of postvocalic /r/ in spontaneous Standard Dutch speech

Leendert Plug
University of Leeds

It has been suggested that speakers of Standard Dutch have a tendency to 'delete' /r/ in postvocalic position, particularly when /t/ or /d/ follows in the same syllable (Kessens, Wester and Strik 1999, Van den Heuvel and Cucchiarini 2001). Plug and Ogden (2003) argue against a categorical view of '/r/-deletion', on the basis of the observation that apparently /r/-less forms may lack a rhotic segment, but still contain correlates of rhoticity distributed across neighbouring segments. They show that in a small corpus of lexical items with and without postvocalic /r/, elicited from four speakers, these non-segmental correlates are found sufficiently widely to maintain contrasts between apparently /r/-less realisations of forms with /r/ and their minimal-pair equivalents without the phoneme. This study assesses the extent to which Plug and Ogden's findings can be generalised to spontaneous speech. It replicates their analysis using data sampled from Corpus Ernestus, which comprises approximately 15 hours of spontaneous conversation produced by 10 pairs of male speakers of Standard Dutch (Ernestus 2000). All lexical items with a stressed vowel followed by /rt/, /rd/, /t/ or /d/ were selected for analysis, resulting in a data set of 635 tokens. All tokens were subjected to auditory and acoustic analysis covering rhyme, vowel and plosive duration, vowel formant structure, and spectral properties of the plosive burst. Measurement results were processed statistically with speaker identity as a random factor and the presence vs absence of the phoneme /r/ as the crucial fixed factor. Results suggest that segmental realisations of /r/ are virtually unattested in spontaneous Standard Dutch spoken by male speakers, and that several of the non-segmental correlates identified by Plug and Ogden are only weakly observed in spontaneous speech. Still, others appear robust, and across the data set there is sufficient evidence of non-segmental rhoticity to maintain that so-called '/r/-deletion' is a gradient phenomenon involving the non-local realisation of phonological contrast (Coleman 2003).

Subglottal resonances and F3 frequency in English alveolar approximant /r/

Mark J. Jones[1] & Rachael-Anne Knight[2]
University of Cambridge[1] & University of York[1], City University London[2]

Acoustic coupling to the subglottal airways (tracheal coupling) introduces a series of pole-zero pairs into the spectrum. Discontinuities in F2 transitions attributable to the second tracheal zero have been documented (e.g. Ishizaka, Matsudaira & Kaneko 1976; Chi & Sonderegger 2007), and have been posited as a quantal acoustic discontinuity underlying the division between front and back vowels (Stevens 2002; Lulich, Bachrach & Malyska 2007; Lulich in press). The research presented here focuses on tracheal coupling and the production of British English alveolar approximant /r/ which has a low F3 due to a sub- or antelingual cavity in the frequency range of the second tracheal pole-zero pair. The location of the second tracheal pole-zero pair is determined by acoustic analysis of spectra of the high front vowel /iː/ for a number of speakers of British English. Interdental coupling is another possible source for extra resonances in /iː/ (Honda et al. in press), but pending further investigation, subglottal coupling is assumed. The location of the second tracheal pole-zero pair is then correlated with the location of F3 in the production of alveolar approximant /r/ from the same subjects, and dynamic changes in the frequency relationships over time of F3 and the second tracheal pole-zero pair are mapped in /riː/ sequences. The results confirm that F3 in /r/ is located close to the second tracheal pole-zero pair for these subjects. Part of the problem of acquiring satisfactory adult-like realisations of approximant /r/ (Klein 1971; Dalston 1972) may therefore lie in fine-tuning the relationship between F3 frequency and the frequency location of the second tracheal pole-zero pair. Implications for Quantal Theory are mixed. On the one hand, the second tracheal pole-zero pair could function as a boundary between [+back] /w/ and [-back] /r/, as it does for vowels, but as with schwa, which also has a formant in the affected region, speakers may be exploiting rather than avoiding the boundary to form a contrast.

Same phonemic sequence, different acoustic pattern and morphological status: A model

Marco A. Piccolino-Boniforti [1], Bogdan Ludusan[2], Sarah Hawkins [1], & Dennis Norris [3]
Dept. of Linguistics, University of Cambridge[1], Dept. of Physical Sciences, "Federico II" University[2], Cognition and Brain Sciences Unit, Medical Research Council[3]

We aim to build a computational model that will help elucidate how humans use acoustic-phonetic detail to understand speech. Most psycholinguistic models of human speech perception focus on using phonological contrasts to identify lexical items. Our focus is on predicting grammatical structure from acoustic-phonetic detail.

The first step is a proof-of-concept model to distinguish between true and pseudo morphological prefixes in English words, as in *discolour*, in which *dis* is a true prefix, and *discover*, in which *dis* is a pseudo-prefix. Both words have the same first four phonemes, /dɪsk/ but linguistic and phonetic analyses show that pronunciations of pseudo prefixes tend to have a weaker rhythmic beat than pronunciations of true prefixes have (Ogden et al. 2000). Concomitant differences in their spectro-temporal fine structure have been documented by Baker (Baker, 2008; Baker et al. 2007a). Heard in sentences or phrases in noise, such words are less intelligible when a syllable of the wrong morphological type replaces the original initial syllable, compared with when a syllable of the right type is spliced into the same position (Baker 2008; Baker et al. 2007b).

The present work uses Baker's original speech corpus and aims to simulate aspects of her observed results. The computational model comprises two main parts. The acoustic signal is first processed within a cochlear model (Patterson et al. 1988; Meddis 1986) that introduces non-linearities in frequency and loudness. The cochlear output is then transformed into an auditory primal sketch (Todd 1994) which simulates perception of amplitude modulation at various temporal resolutions within the auditory system. This representation identifies successive acoustic events in the signal and their so-called relative prominence, a measure that combines amplitude and duration, and has been used to compare speech rhythm of various languages (Lee & Todd 2004).

In the second stage of the present model, the output of the auditory primal sketch is input to a classifier. Two classifiers are compared, the popular Support Vector Machine (SVM, Vapnik 1995), and the Relevance Vector Machine (RVM, Tipping 2001). In order to assign category membership to a novel sample, both classifiers rely on just a subset of the training samples, thus achieving what is termed a sparse representation of a category. While in the SVM the subset of samples represents category boundaries, in the RVM it represents category prototypes. Prototypes are interesting for their relevance to the perceptual magnet effect and perceptual plasticity. The RVM also displays other desirable properties, including greater sparsity, the output of probabilities rather than sharp category decisions, and the possibility of incorporating prior knowledge, such as word frequency.

The present work reports simulations that compare: 1) RVM vs SVM classification accuracy; 2) the relative merits of the cochlear model and auditory primal sketch as opposed to more standard, energy-based feature vectors. In each simulation, the model was trained on 800 samples and tested on 200. For the RVM vs SVM simulation, the samples were Baker's manually-measured segment durations. For the remaining simulations, the samples were prominence scores, obtained as described above.

Results show that both RVM and SVM assign the data to the correct true vs pseudo morphological category at well above chance (area under the curve (AUC): RVM 0.910; SVM 0.909; chance 0.500). The RVM obtains a much sparser representation (number of training samples used for category decisions: RVM 45; SVM 339). The cochlear model improves the accuracy of the auditory primal sketch (AUC: without cochlear model 0.719; with cochlear model 0.764). However, classification is more accurate using energy vectors than the auditory primal sketch (AUC: energy 0.919; auditory primal sketch 0.764). Nevertheless, despite poorer performance in this task the auditory primal sketch provides a perceptually motivated, automatic segmentation of the acoustic signal. It is therefore preferred for future development of the model, where it can be associated with complementary representations of fine spectral structure.

Segmentation cues in spontaneous and read speech

Laurence White, Lukas Wiget, Olesya Rauch & Sven L. Mattys
Department of Experimental Psychology, University of Bristol

Background. Speech segmentation research asks how listeners locate word boundaries in the ongoing speech stream. Previous work has identified multiple cues (lexical, segmental, prosodic) which affect perception of boundary placement (e.g. Cutler & Norris, 1988; Norris, McQueen & Cutler, 1995; Quené, 1992). Such studies have almost all been based on careful read speech; however, the realisation of boundary cues may be modulated by the interactive and contextualized nature of spontaneous speech. In particular, degree of articulatory effort – hyperarticulation vs hypoarticulation – has been held to vary as a function of communicative and situational demands (e.g. Lindblom, 1990). Cues that are highly salient due to hyperarticulation in non-contextualised speech may be reduced where lexical content is predictable, either as a result of contextual expectation or as a result of repetition of words or phrases.

Method. We report development of speech corpora designed to examine the production of segmentation cues in natural conversational speech. Parallel corpora of English spontaneous and read speech allow us to: (1) compare the realisation of word-boundary relevant information in the two speech styles; and (2) test listeners' utilisation of the segmentation cues present in spontaneous speech. To elicit spontaneous speech whilst controlling boundary-relevant properties, we adapted the Edinburgh Map Task (Anderson, Bader et al., 1991), in which speakers interact conversationally regarding a route around landmarks on a map. In our task, landmark names were one-word or two-word phrases, all paired with similar phrases contrasting in terms of potential word boundary cues: e.g. near-homophonous phrase pairs such as *great anchor* vs *grey tanker;* matched non-ambiguous phrase pairs, such as *bright anchor* vs *dry tanker*. Ten speakers of standard Southern British English were recorded in pairs, and speakers were familiarised with the landmarks in an initial training phase until they could reliably name them without text. To compare the realisation of cues between spontaneous and read speech, all map description utterances containing landmarks were orthographically transcribed and a subset re-recorded later as read speech by the original speakers.

Corpus analyses. A wide range of phonetic analyses are in progress. Here we report results relating to two segmentation-relevant durational phenomena: contrastive rhythm (i.e. stressed syllable lengthening) and word-initial lengthening. Articulation rate was consistent between speaking styles, thus allowing direct comparison of durational effects. The contrastive rhythm analysis indicated that variation in vocalic interval duration is greater in spontaneous than read speech, thus potentially providing stronger metrical segmentation cues. Word-initial lengthening was robustly observed in both read and spontaneous speech, and whether or not phrases were ambiguous, but the contrast between consonant duration in initial and final positions was greater for ambiguous tokens (e.g. *great anchor* vs *grey tanker*) in read than in spontaneous speech. This suggestion of relative hypoarticulation in spontaneous speech was supported by a perceptual experiment examining the effect of repetition on the ambiguity of phrases like *great anchor* and *grey tanker.* We found that, in spontaneous speech but not in read speech, such phrases were more ambiguous when repeated than on first occurrence. Overall, results suggest a shift in cue dominance in spontaneous speech compared with read speech, from localised cues like initial lengthening to prosodic cues like stress.

An investigation of word juncture development in one typically developing child

Sarah C. Bryan, Sara Howard & Mick Perkins
Department of Human Communication Sciences, The University of Sheffield

Theories of speech development have traditionally focused on individual phones and their interactions within the word. However, research in the 1980s found evidence of phonetic interactions occurring between words in children's early multi-word utterances (Matthei, 1989; Stemberger, 1988; Donahue, 1986). These findings have been followed up in recent investigations into the development of connected speech processes (CSPs), which occur at word junctures and are well-documented in adult speech (Shockey, 2003; Cruttenden, 2001). These include assimilation, elision and liaison. Assimilation, elision and liaison of [j] and [w] appear to emerge relatively early, whereas /r/ liaison emerges later and more suddenly, although there exist considerable individual differences (Thompson and Howard, 2007; Newton and Wells, 2002; Newton and Wells, 1999). These findings raise questions concerning the extent to which CSPs are coarticulatory phonetic phenomena or learned phonological behaviours. It is also possible that the individual variability observed is a further reflection of the analytic and holistic language learning strategies originally proposed by Peters (1977).

The current study investigated the development of CSPs in a typically developing child from age two to four years. Samples of audio data were selected at monthly intervals over 25 months from the dense database of child language compiled by Theakston et al. (2001), thus providing a more longitudinal data sample than has previously been possible. Utterances which contained potential CSP sites were analysed using narrow phonetic transcription, in order to identify both the specific behaviours occurring at CSP sites and overall developmental trends. Assimilation was found to emerge early as predicted, although assimilation sites were realised with greater variability at a later age than previously found. As previously reported, /r/ liaison emerged suddenly, several months after the emergence of potential sites. In addition, some instances of open juncture at liaison sites could be explained in terms of interactional context, for instance repetitions in response to requests for clarification. A range of non-adult idiosyncratic phonetic behaviours was also observed at CSP sites. These findings support the general developmental trends and individual variability observed in previous research, in spite of considerable methodological differences and the more longitudinal nature of the current data sample. This study supports suggestions from previous studies that interactions may exist between the development of CSPs and syntax. Future research aims to investigate this further using a denser sample from the same database.

Discrimination of foreign accents of English in the forensic context

Allen Hirson
City University London

The question of the confusability of foreign accents of English presented recently in the forensic context. The assailant in a case of multiple rape was identified by the injured parties as having a Polish accent of English but the suspect was a native speaker of Palestinian Arabic. Accent disguise was discounted owing to the speaker's competence in English. The research question is the extent to which one foreign accent might be confused with another.

Based upon the known immigrant groups in the area where the rapes took place, two Slavic languages (Russian and Czech) were used as perceptual distracters for the Polish accent, and two Semitic languages (Egyptian Arabic and Tigrigna) served as distracters for the Palestinian Arabic accent. Three unrelated languages (Romanian, Farsi and Turkish) served as additional foils. Read samples of these nine foreign accent samples were, with the exception of the Palestinian sample, from the Speech Accent Archive (http://accent.gmu.edu/, 2009). The Palestinian sample was recorded from the suspect in prison.

A perceptual experiment involved playing the accent samples to a group of native British English females with normal hearing none of whom had significant experience of Slavic or Semitic languages. The listeners identified where in the world each accent originated, and the confidence of their judgements. Previous research on the perception of speaker nativity highlights the significance of suprasegmental features (e.g. Magen 1998; de Mareüil & Vieru-Dimulescu, 2006). However, other research (e.g. by Major, 1987) emphasizes the role of segmental detail in influencing the perception of foreign accented speech.

Present findings analysed using a form of association statistics known as correspondence analysis (Clausen, 1998) revealed a remarkably organised pattern of listeners' judgements - despite the low levels of confidence rated by the listeners themselves. Since prosody is somewhat disrupted owing to the L2 reading competence, it is suggested that these listener judgements must be accounted for by segmental detail contained in the samples. The observed errors were also of interest. For example, the Russian-accented sample was sometimes mis-identified as 'Dutch' (reflecting anecdotal accounts of the same confusion), and the Romanian-accented sample was frequently and correctly identified as being derived from a Romance language. These relationships of listener judgements to target accent is graphically displayed by the association statistics. The idea of determining 'distances' between accents has previously been applied to Norwegian dialects (Heeringa, Johnson & Gooskens, 2009). It now shows some promise when applied to foreign accents of English.

The Influence of Context on Listeners' Perceptions of Distress and Linguistic Content in Forensic Audio Recordings

Lisa Roberts, Peter French, & Philip Harrison
University of York & JP French Associates, York

The aim of the experiment was to assess the influence of context on listeners' perceptions of speech, vocalisations and screams produced by victims undergoing violent attack. Stimulus data were drawn from edited sound files drawn from real forensic audio recordings.

Listeners - a group of experienced forensic phoneticians and a group of postgraduate student phoneticians without casework experience - were presented with brief stimulus sounds and asked to:

(a) categorise them in terms of a 4-way classification corresponding to the degree of distress they perceived the stimulus as representing;

(b) rate them on a 5-point scale specifying the degree to which they perceived the stimulus material as having linguistic content.

Listeners heard the stimulus sounds under two conditions. Under condition one the sound was heard in isolation, stripped of its original context. Under condition two the listeners heard not only the sound in question, but what had preceded it and what followed it. They were also given background information concerning the circumstances of the recorded attack.

Findings indicate:

(a) stimulus sounds were generally rated to reflect higher degrees of distress when heard in isolation;

(b) stimulus sounds were generally rated as having lower levels of linguistic content when heard in isolation.

(c) ratings of the experienced forensic phoneticians were less prone to change across the conditions. However, where ratings were changed, the same general change-trend was observed.

Insofar as forensic phoneticians are regularly asked to transcribe recordings of attacks, these findings underscore the fact that the ascription of linguistic content to a victim's brief utterance, scream or vocalisation is unlikely to be achieved by a phonetic consideration of the internal properties of the sound itself. As, for example, Fraser (2003) has suggested, higher order information – including sequential context and background story – plays a pivotal part. Similarly, the impression of distress is only partly conveyed by the sound itself; contextual information is an important factor.

Voice Similarity and the Telephone

Toby Hudson, Kirsty McDougall & Francis Nolan
University of Cambridge

Although the topic of voice quality has been widely discussed, relatively little is known about what makes voices sound similar or different to listeners. As well as being of theoretical interest in phonetics, progress on this is crucial for optimising earwitness evidence in forensic phonetics. In the latter context the effect of the telephone on how voices are perceived is also important. This paper presents results from two relevant experiments on voice similarity where the salient linguistic variable of accent has been excluded.

In the first experiment, subjects rated paired voice samples on a difference scale. The samples were taken from fifteen 18-25 year old speakers of Standard Southern British English included in an accent-controlled database. For each pairing of speakers (including same-same pairs) 20 listeners heard two short spontaneous speech samples taken from a telephone interaction, and were asked to rate the distance between the two voices on a scale of 1 to 9. The speech samples had been recorded simultaneously in both studio and telephone quality and were heard in 'studio only', 'telephone only', and 'mixed (telephone and studio)' pairs. Results show that, as might be expected, samples heard in the band-limited telephone condition are heard as more similar; but the 'mixed' condition yielded less predictable results. In order to probe what underlies the perceptual ratings, correlations were computed between the ratings (on studio samples) and a number of salient acoustic characteristics of the speakers.

The second experiment investigated the effect of the telephone on (mock) earwitness voice identification, using a voice parade constructed in accordance with guidelines current in England and Wales. Multidimensional scaling of the similarity ratings from the first experiment was used to select 9 well-matched speakers from the 15 rated. For 5 speakers out of the 9, a sample of speech not overlapping with that used in the parades was selected as the material for listeners' initial exposure to a 'target' speaker. One hundred listeners took part in the experiment, 25 in each of 4 conditions: exposure and parade both at studio quality, exposure and parade both at telephone quality, studio exposure/telephone parade, and telephone exposure/studio parade. Within each condition 5 listeners were allocated to each of the 5 target speakers. One week after exposure to a target speaker the listeners attended a (closed-set) voice parade and were asked to identify the voice they had heard the week before. The results show that identification performance holds up reasonably well under the telephone condition. The mixed conditions, however, show a striking effect depending on whether the telephone quality was applied to the target presentation or the parade, and an explanation for this will be proposed. This finding offers a clear practical message for the construction of voice parades.

An acoustic analysis of syllable-initial /l/ in Glasgow Asian

Jane Stuart-Smith[1], Claire Timmins[2] & Farhana Alam[1]
University of Glasgow[1]; Queen Margaret University Edinburgh[2]

The increase in minority ethnic groups in the UK, and widespread informal acknowledgement of ethnic accents within and beyond communities, has led to an increasing interest in ethnicity and accent (e.g. Kerswill et al 2008; Sharma and Sankaran 2009), but there is still little phonetic research on regional ethnic British accents (Foulkes and Docherty 1999, but see e.g. Heselwood and McChrystal 2000; Hirson and Sohail 2007). The main ethnic group in Glasgow is from the Indian subcontinent, and the 'Glasgow Asian accent' is accepted to the extent that it is even represented (stereotypically) in a character in a local TV comedy show. A recent study based on auditory analysis of two complementary datasets identified specific accent features for Glasgow Asian, and also suggested that these features vary according to social/cultural practices (Lambert et al 2007).

This paper presents the results of an acoustic analysis of syllable-initial /l/ in word-initial and word-medial position (e.g. *letter, daily*) in the same data. The auditory analysis identified clearer realizations of /l/ in Glasgow Asian speakers; Glaswegian typically shows dark /l/ in all positions in the word. The data were taken from two small-scale Studies: Study One consists of read passages by 6 Glasgow Asians and 4 Glasgow non-Asians. Study Two comprises spontaneous speech from 7 Glasgow Asian girls from an ethnographic study of language and social practices, ranging from 'traditional' to 'modern', with regard to their affiliation to religious/cultural heritage.

Following Carter and Local (2007), the waveforms for all possible tokens were labelled to identify the transition into the lateral, lateral steady state, transition out of the lateral, and the vowel. Using the boundaries for these phases the first four formants were taken at 10 equally-spaced points across the lateral-vowel portion. Here we focus on some acoustic correlates to clearness/darkness differences in /l/ (Carter and Local 2007; cf also Hawkins and Nguyen 2004): frequency of F2 at steady state; trajectory of F2 from lateral into the vowel; and relative durations of the phases of lateral and vowel.

The results for the steady state show significant differences according to ethnicity in Study One, and identity and social practices in Study Two: Glasgow Asian speakers show higher F2 than Glasgow non-Asian, and the most traditional Glasgow Asian girl also show the highest F2. However, whilst Glasgow non-Asian speakers show expected low F2 values (average 943Hz, male speakers) reflecting dark /l/, Glasgow Asian speakers show values which are still within the range of dark /l/s in other dialects of English, e.g. Leeds (average 1092Hz, male speakers; see Carter and Local 2007; Recasens and Espinosa 2005). Normalized durations also show differences according to ethnicity and identity: Glasgow Asian speakers show tendencies for shorter transitions into /l/, longer steady state and significantly shorter transitions into the vowel and longer following vowels; again the most traditional girl from Study Two shows a similar pattern. But the duration measures also show good separation between the girls according to social practices. Finally, inspection of the time-normalized F2 tracks across the phases of lateral-vowel show different dynamic patterns according to ethnicity, and also very close patterning between formant trajectory and individuals' reported engagement with ethnic networks in Study One, and social practices in Study Two.

These findings confirm clearer realizations of /l/ in Glasgow Asian, but also show how relative points along the clearness/darkness continuum may be exploited within a regional dialect. They also demonstrate the degree to which fine-grained acoustic differences may relate to the construction of locally-salient ethnic identities.

An acoustic and articulatory investigation of high vowels in Scottish English children and adults

Natalia Zharkova[1] & Robin Lickley[2]
Speech Science Research Centre[1], Queen Margaret University[2]

The present study addresses acoustic and articulatory properties of high tense vowels in Standard Scottish English (SSE) child and adult speakers. High vowels /i/ and /u/ in adult SSE are closer together in the acoustic and articulatory space than these vowels in Southern British English. In children acquiring Scottish English phonology, realisations of /u/ tend to be at first more front and variable, gradually changing to more back productions (Scobbie, Gordeeva & Matthews 2007). There is no information on how long this process takes, nor is there articulatory evidence of how children develop a consistent distinction between these vowels.

A recently collected database of SSE child and adult productions (Zharkova 2009) has provided impressionistic evidence that these two vowels sound very similar to each other in some 6 to 8 year-old children. The present study analyses acoustic and articulatory data from this database, in order to find out to what extent children who have generally acquired the phonemic system are different from adults in how they differentiate between /i/ and /u/. The objectives of the study are: 1) to establish whether /i/ and /u/ are significantly different from each other in children, as well as in adults; 2) to identify and explain the nature of any age-related differences, using complementary information from acoustic and articulatory data.

The database consists of synchronised recordings of the acoustic signal and lingual articulations, the latter obtained using ultrasound tongue imaging. Ultrasound provides information about the shape of most of the midsagittal tongue contour, including the root (e.g. Stone 2005). The speakers selected for this study are six female adults and six children aged between 6 and 8 years. Ten repetitions of each target vowel are produced within a /s/-vowel syllable in a carrier phrase.

The data analyses are as follows. In each token, F2 is calculated at the vowel midpoint. Separately for children and adults, F2 values for /i/ and /u/ are compared, using T-tests with Bonferroni adjustment. The tongue surface contour is traced at the vowel midpoint in each token, and defined in terms of x-y coordinates. Distances between tongue curves are used to compare /i/ and /u/ tongue contours. Separate Anovas for children and adults establish whether the tongue position for /i/ is significantly different from the tongue position for /u/. A correlation is performed between acoustic and articulatory results. In addition to the statistical analyses, qualitative examination of individual variation in tongue shapes and spectrograms will be used to explain the nature of any age-related differences in realising the vowel distinction. The results will inform theories of speech development and practical descriptions of English phonetics.

Production and perception of word boundaries across accents

Rachel Smith[1] & Rachael-Anne Knight[2]
University of Glasgow[1], City University London[2]

It has been claimed (Abercrombie 1979) that cues to word boundaries differ across accents of English. If this is the case, it might be more difficult to segment the speech stream when listening to an accent different to one's own. Segmentation of speech might also be easier in an accent that is heavily represented in the national broadcast media, (even by people who do not speak with that accent), compared to an accent that receives more limited media representation (e.g. Adank et al., 2009).

Two experiments were conducted. The production experiment aimed at identifying the acoustic cues to word boundaries in Standard Southern British English and Glaswegian. Speakers recorded pairs of sentences containing phrases that varied according to the position of a word boundary, (e.g. 'he diced them vs. he'd iced them'). The durations of vowels, consonants and syllables was compared between the items in each pair. There were some similarities between the cues in both accents. For example, consonants are longer in onset than coda position (e.g. /d/ is longer in 'he **d**iced' than 'he'**d** iced'). However, the difference between the two /d/s is significantly larger in SSBE. In general certain aspects of word boundaries seem acoustically less distinct in Glaswegian than SSBE.

The perception experiment addressed how successful word segmentation and identification are when listening to speech in a familiar vs. an unfamiliar accent. In the pre-test, the sentence pairs from the production experiment were played to listeners in noise, and they wrote down what they heard. Then, in the exposure phase listeners heard the same sentences in good listening conditions, in a context that made clear which phrase was intended (e.g. 'He wanted the carrots to cook fast, so he diced them'). Finally, the post-test stage was a repeat of the pre-test. Half the listeners were exposed to their own accent, and half to the other accent. Improvement between pre- and post-test was calculated for the percentage of words correctly identified, and for segmental variables such as correct identification of onset and coda consonants.

Listeners always comprehended speech in noise better when they were listening to their own accent, suggesting that a simple account based on media representation needs refining. However, Glaswegian listeners tended to *improve* more between pre- and post-test with both accents than SSBE listeners did, and both groups of listeners improved more with SSBE than Glaswegian. In part this may be due to clearer word boundary cues in SSBE, meaning that all listeners can learn to use them more easily. However, Glaswegian listeners might also be *underlyingly* more familiar with SSBE. Despite their comprehension being impaired upon a first encounter with the accent, they recover and improve more substantially after exposure. This could suggest that media exposure is still an important factor in speech perception, but has a more subtle effect than previously suggested.

Updating the Scottish accent map: preliminary formant data from the VOYS corpus

Catherine Dickie[1], Christoph Draxler[2], Felix Schaeffler[1], & Klaus Jänsch[2]

Speech Science Research Centre, Queen Margaret University, Edinburgh[1], Institute of Phonetics and Speech Processing, Ludwig Maximillian University, Munich[2]

The VOYS (Voices of Young Scots) corpus is a speech database of Scottish English spoken by adolescent speakers (aged 13 to 18). The recordings for this corpus were performed in Scottish secondary schools between autumn 2008 and autumn 2009, using the web-based WikiSpeech system (Draxler & Jänsch, 2008). All recordings comprise two channels (Beyerdynamic Opus 54.16/3 headset microphone and AT3031 table microphone), digitally recorded with 22 kHz and 16 bit.

Data collection has been completed in seven secondary schools in six Scottish locations: Inverness, Aberdeen, Edinburgh (2 schools), Jedburgh, Ayr, and Dumfries. Participating schools were selected by the socio-economic factor of "free school meals". Schools were chosen that were close to the local average for this factor and thus likely to be representative for the area.

Currently, the database contains 175 speakers, more than 16800 utterances and a total recording duration of approx. 30 hours.

The speech material consists of

1. scripted speech

    a) read words and sentences targeting certain sociophonetic aspects of Scottish English

    b) read digits, numbers, currency, date and time expressions, and phonetically rich sentences (this part of the corpus is speech-dat compatible (Höge et al., 1999);

    c) a read story (the Dog and Duck story; Brown & Docherty, 1995)

2. unscripted speech

    This part consists of spontaneous speech, elicited by questions like "please describe your way to school" or descriptions of pictures showing dramatic events.

For the analysis presented here, the words and sentences from category 1a were used. Currently there are about 7500 items in this category. These items were automatically segmented and labelled, using the Munich Automatic Segmentation System (MAUS; Schiel, 2004). We will present F1 and F2 values for nine Scottish English monophthongs as they are realised in each of the five locations, and discuss the implications for the role of vowel variation in distinguishing regional varieties in Scotland.

A first release of the VOYS corpus with these recordings will be made available in early to mid 2010.

Voice Onset Time and the Scottish Vowel Length Rule along the Scottish-English border

Dominic Watt, Carmen Llamas, Jennifer Nycz & Damien Hall
University of York

This paper reports on two durational properties of Scottish and northern English English, Voice Onset Time (VOT) and the Scottish Vowel Length Rule (SVLR), in 4 towns along the Scottish-English border: Gretna (west) and Eyemouth (east) in Scotland, Carlisle (west) and Berwick-upon-Tweed (east) in England. These towns form pairs in which the towns are nine miles apart and separated by the border. Though both VOT and SVLR have been examined in detail in many locations in Scotland and (to a much lesser extent) northern England, no previous study has made direct comparisons both across and along the border.

According to Catford (1988) (see also references in Watt & Yurkova 2007), word-initial voiceless stops are typically unaspirated in Scotland and northern England, but aspirated in other British varieties. The SVLR, as described by Aitken (1981), affects certain vowels such that they are long in open syllables and preceding the voiced fricatives /v ð z ʒ/, /ɹ/, and morpheme boundaries, and short elsewhere (cf. the conditioning environments for the more general Voicing Effect; Chen 1970). Our study investigates the status of each of these features in the four localities, whether there is change over apparent time, and if so, whether the four varieties are changing in different ways.

VOT was measured for tokens of initial /p t k/ from wordlist readings. For SVLR, duration measurements were taken for tokens of relevant vowels in both triggering and non-triggering environments.

Our study yields several novel findings for each variable. We confirm that Scottish speakers have shorter VOTs than the English ones, yet in all four communities there is evidence of change, and in the same direction: young speakers in all four towns have longer VOTs than older ones. However, the young Scottish speakers still have shorter VOT than the old English speakers. In addition, there is parallel change among males at each end of the border (Berwick’s VOT is changing at same rate as Eyemouth’s, and Carlisle’s at the same rate as Gretna’s), though the rates of change in the eastern and western pairs of towns are not the same. Regarding SVLR, we find that young speakers in Berwick, Eyemouth and Gretna (those of our communities where SVLR has previously been reported) show less evidence of this feature than their older counterparts; the young speakers seem to be converging on a more general Voicing Effect.

For both these features, then, we find that younger speakers appear to be moving in the same direction, possibly towards the norms typically associated with southern English varieties. However, our study also reveals the value of challenging the assumption that the presence of the border has the same effect on language use all the way along its length, since it seems that for VOT the east-west dimension between our sites plays at least as important a role as the cross-border dimension (as the sites across the border from one another are changing at the same rate as each other, but not at the same rate as the sites at the other end of the border). For SVLR, there seems to be more convergence both along and across the border.

ABSTRACTS

Wednesday 31st March 2010

Negotiating towards a next turn: Phonetic resources for ‘doing the same’

Rein Ove Sikveland,
University of York

When we continue talking beyond a possibly complete turn, the continuation will be shaped and understood in relation to the previous talk; this is how we achieve coherent meaning (Heritage 1984). Joining and separating units of talk may involve different levels of phonetic detail, as has been shown in recent work on talk in interaction (e.g. Couper-Kuhlen 2003, Curl et al. 2006, Local 2004, Local & Walker 2004). In this presentation I will show how interactants use resources for continuing on the same action, or ‘doing the same’, and how phonetic detail is a significant part of these resources.

The material presented will be based on a 120 minute audiovisual collection of spontaneous dialogues in Norwegian. The material has been analysed using a method that combines Conversation Analysis (CA) and linguistic analysis. This method puts the participants and their displayed orientations at the centre of analysis, and not the analyst’s (Local & Walker 2005).

A reason a main speaker might have for ‘doing the same’ as before is to maintain (i.e. not changing) an action across turn units. As we will see, the co-participant (or recipient) has similar resources to *treat* preceding turn units as ‘doing the same’. I will show how such a practice involves a range of phonetic features, depending on the phonetics of previous turns (i.e. the turns that one is treating as ‘the same’). This process is also relevant in terms of how participants distinguish between passing and taking up on an opportunity to talk next (cf. Curl et al. 2006).

I will focus on sequences of talk where ‘doing the same’ is relevant for the negotiation of speaker change; a main speaker’s talk may have come to a point where it is relevant for another to talk, but this change will be established in a stepwise manner. I will argue that whether or not interactants use this practice has consequences for how the conversation proceeds.

With this kind of research I aim to account for aspects of phonetic variability in talk in interaction, by showing how much phonetic variation is allowed when shaping an action, or meaning. Also, it forms a detailed and rich understanding of how interactional meaning is an online, moment-by-moment achievement involving speakers *and* their co-participants (Goodwin, 1979). My presentation will provide an example of how interactional meaning does not inhere to single turns of talk, but rather develops through several turns of talk, as a shared achievement between those who interact.

Say it again: the phonetics of lexical repetition in Welsh

Paul Carter
Bangor University

Some recent work on the phonetics of talk-in interaction has explored aspects of the phonetic shapes associated with lexical repetition in several varieties of English (as it is related to social actions within the structure of conversation). For example, Curl, Local and Walker (2006) investigated what they refer to as ‘doubles’; that is, repetitions used to close sequences of talk; Curl (2004, 2005) has also reported on the phonetic details of other-initiated repair.

In this paper I will test some of the claims made for English against a corpus of conversational data from another language: Welsh. The Welsh data set comes from a corpus of 40 hours of conversation in Welsh (with some code-switching into English). Preliminary investigations provide evidence which broadly supports the analyses Curl, Local and Walker provided for English.

The doubles examined so far share with English the phonetic shape in which both instances have the same stress pattern and a falling pitch; the highest pitch of the second instance is lower than the highest pitch of the first and there is minimal variation in articulatory characteristics. But it is not yet clear that other observations made by Curl, Local and Walker also hold true for this corpus of Welsh data: the second instance may not necessarily be shorter in duration, nor is there necessarily an equivalence in loudness between the two instances.

In the case of other-initiated repair, however, there seems to be remarkable agreement with the data from English. For turns produced in the clear (i.e. without overlapping speech), re-doings which are interactionally fitted to the preceding turns are produced with what Curl called ‘upgraded’ phonetics: typically louder, with an expanded pitch range, longer duration and with changes in articulation. Re-doings which are disjunct from the preceding turns are produced with non-upgraded phonetics: typically quieter, without an expanded pitch range, shorter duration and minimal articulatory change. Curl also demonstrated for English that turns produced in overlap sometimes share the characteristics of turns produced in the clear (i.e. upgraded phonetics for fitted turns and non-upgraded phonetics for disjunct turns), and sometimes share the non-upgraded phonetic characteristics of the disjunct in-the-clear turns even if the turn is fitted: the difference between these two speech-in-overlap patterns is accounted for by reference to the detail of the interactional sequence of turns. Again, I observe the same patterns in Welsh.

I will also outline some directions for future work, including a comparison with a new corpus of conversational data from Welsh-Spanish bilinguals.

Speech rate as a resource for turn competition in overlapping speech

Emina Kurtić
Department of Human Communication Sciences / Department of Computer Science University of Sheffield

The exchange of speaking turns between participants in conversations mostly happens smoothly - each turn is allocated to one speaker upon the current speaker's turn end. This ensures smooth turn transitions with minimal gap between speakers' turns and also with minimal speaker overlaps. However, occasionally situations arise in which several speakers claim the turn at the same time. This leads to turn-competition, and it often results in overlapping speech.

Most previous research on overlap has agreed that some instances of overlapping speech are turn competitive while others are not. Raised pitch and loudness have been identified as phonetic features that speakers use for turn competition and also orient to as turn competitive (Kurtic, Brown & Wells, in press). However, little is known about the role of speech rate as a turn competitive resource.

In extension of this previous work on prosodic resources for turn competition this paper presents a study on how participants use speech rate to achieve and respond to turn competition in overlap. Our analysis is based on a set of overlaps extracted from the ICSI Meeting Corpus, which contains transcribed recordings of face-to-face, multi-party research meetings. Sequences of overlapping talk were identified in the data and annotated as competitive or non-competitive based on the observable orientation of conversation participants. Speech rate is measured at the overlap onset, within the overlapping section and at the overlap resolution point of each overlap instance. We compare speech rate in these local contexts and also across speakers involved in overlap. By contrasting the speech rate of competitive and non-competitive overlaps we are able to describe how this feature is used by participants when competing for the turn in overlap. We also describe how overlapped speakers use speech rate to respond to competitive and non-competitive overlaps.

‘Intensifying emphasis’ in conversation

Richard Ogden
Department of Language & Linguistic Science, CASLC, University of York

Intensifying emphasis (Niebuhr, ms, Niebuhr & Kohler 2007) is the name given to a linguistic and phonetic-prosodic construction whose function seems to be to mark items out as in some way ‘exceptional’. Here follows an example: the speaker is explaining how she arrived in an unknown city as a baseball game was starting, and she got stuck in traffic:

‖ I didn’t even know where I was going | you know | I was just following the directions

‖ well the traffic ‖ I was beside myself ‖

Here, the speaker produces ‘traffic’ as [\tɹ̝̊ːæːfɪk]. ‘Well the’ lasts 250ms (8 syll/sec); ‘traffic’ lasts 710ms (2.8 syll/sec): there is thus a marked decrease in tempo in the ‘intensified’ speech. The closure for the plosive lasts 75ms, and the VOT 165ms. The f0 falls almost 8 semitones from the peak in the first syllable to the end of the second syllable.

This paper presents results from a study of naturally occurring instances of intensifying emphasis. The data come from 54 speakers, and approximately 540 minutes of unscripted speech from 27 calls from the CallHome corpus. In all, there are approximately 120 examples of intensifying emphasis in a set of 85 extracts. Phonetically, the data exhibit a number of features:

- there is a falling pitch contour on the accented (intensified) item;
- the articulations associated with the intensified item are tense: e.g. aspiration portions of voiceless plosives are long; there are long periods of voicing and closure for voiced plosives; some fricatives are produced with a plosive onset; vowels start with glottal stops;
- the intensified items are frequently preceded by periods of closure or a short gap in production;
- the tempo of the speech is faster before the intensified item, and considerably slower on and after the intensified item.

In addition, there is a cluster of other linguistic features which cohere with intensification. The lexical form of intensified items is most likely to be a modifier or an adjective; nouns are also frequent; verbs and quantifiers are also found but less commonly. Intensification often occurs alongside ‘extreme case formulations’, i.e. formulations where the speaker makes a strong (possibly exaggerated) assessment.

A consideration of the sequential placement of intensification and its treatment by coparticipants shows that it is used for a cluster of related functions:

- to portray something as ‘tellable’
- to upgrade an assessment that has already been made or which is already implied
- to handle epistemic issues, such as which speaker has a stronger claim to make an assessment

Intensifying emphasis is an example of a ‘prosodic construction’: a cluster of phonetic and other linguistic features which have a recurrent meaning. Such constructions raise interesting questions about the relation between phonetic variability and detail, sequential placement, and communicative function.

A multisystemic model of rhythm development: Phonological and prosodic factors

Brechtje Post[1], Elinor Payne[2], Lluïsa Astruc[3], Pilar Prieto[4,5], & Maria del Mar Vanrell[4]
RCEAL, University of Cambridge, and Jesus College[1], Phonetics Laboratory, University of Oxford, and St Hilda's College[2], Department of Languages, The Open University[3], Departament de Traducció i Ciències del Llenguatge, Universitat Pompeu Fabra[4], ICREA, Institució Catalana de Recerca i Estudis Avançats[5]

The characteristic rhythm of a language – traditionally referred to in terms of stress- and syllable-timing – has been claimed to emerge from various phonological properties, especially vowel reduction and syllable complexity (e.g. Bertinetto 1981, Dasher and Bolinger 1982, Roach 1982, Dauer 1983; cf. Prieto et al. submitted). For a child, learning how to produce speech rhythm must be a complex task; it not only needs to develop sufficient motor control to vary pitch, duration, loudness and spectral properties to achieve an appropriate rhythmic pattern, but the child also needs to acquire phonological features like vowel reduction, syllable and segment timing, syllable structure, and stress assignment. Acquiring rhythm must therefore, of necessity, be closely intertwined with the acquisition of other aspects of the language system. This implies that, while rhythmic development may start early (e.g. Nazzi et al 1998), it will not be complete until all of the key segmental and suprasegmental properties are fully acquired, which potentially encompasses the entire period of phonological development until approximately age 9 (Ruscello 2003).

The few existing developmental studies of rhythm production appear to support this view. While some language-specific prosodic properties begin to emerge in production by the end of the first year (Boysson-Bardies and Vihman 1991), a study comparing French and English children's productions showed that the English children do not yet master their native rhythm at 4 (Grabe et al 1999). Grabe and colleagues also found that the children produced more 'even' timing than their parents (also e.g. Lléo and Kehoe 2002). They argued that acquiring the rhythm of a stress-timed language like English is more complex, since it requires the child to move away from the 'even' timing it sets out with, involving the acquisition of language-specific properties like vowel reduction, complex syllable structures, and variable accent placement.

In an earlier study, we found that rhythm (measured on a variety of metrics), does indeed change with child age (2-, 4-, and 6-year olds, 36 children in total; Payne et al. submitted)). However, although the children's speech was more 'vocalic' (higher %V) than the adult targets (produced by their mothers), and it had lower variability in the duration of vocalic intervals, consonantal interval variability showed the opposite pattern, with higher variability for younger children, which decreased with age. This does not support the hypothesis that child speech starts out more even-timed across the board. We also observed cross-linguistic differences in the developmental paths for English on the one hand (stress-timed), and Spanish (syllable-timed) on the other, which grouped with Catalan (mixed/intermediate rhythm class). The most striking difference was that even at 6, the English children were still quite un-target-like.

Using the same data, we investigate to what extent these findings can be attributed to cross-linguistic developmental differences emerging from other systemic properties like syllable structure, stress placement, phrasing and segment inventory. We hypothesise that rhythmic differences emerge in parallel with the acquisition of phonology (especially syllable structure). Hence, cross-linguistic rhythmic differences in adult speech should already be apparent at age 2, but become stronger with age, reflecting phonological and prosodic differences. Initial findings show that, although there are indeed ambient effects of syllable structure at age 2, syllable complexity cannot fully explain the findings. Factors such as stress placement and phrasing, which are some of the properties which are not yet fully acquired at 6, also play a role, supporting a more refined, multi-systemic model of rhythmic development.

Surprise? Surprise! Cue interaction in rises and falls with different functions

Toby Hudson[1], Brechtje Post[1], Iwo Bohr[1], Francis Nolan[2] & Emmanuel Stamatakis[3]
Research Centre for English and Applied Linguistics[1], Department of Linguistics[2], Division of Anaesthesia, School of Clinical Medicine[3], University of Cambridge;

Unlike in segmental phonology, there is no minimal pair test for intonational units; that is, meaning is not expressed in a straightforward one-to-one fashion with variation (Ladd 1996, Martinet 1962). Intonational variation can be categorical, for instance when a rise marks a question (without systactic marking) in Standard Southern British English (SSBE), as opposed to a fall marking a statement. Variation can also be gradient, for instance when scalar variation in the extent of the rise is used to signal paralinguistic information, such as emotion (higher pitch peak for surprise), or when it results from the Lombard effect (increased pitch, along with intensity, against background noise). However, a gradient ‘phonetic’ change (e.g. increase in pitch peak) can itself imply a categorical, linguistic distinction of function (higher peak for question than for continuation rise). Thus the situation is complex, and detail plays an important role (Nolan 1999).

The main objective of the present research is to provide evidence for the hypothesis that categorical, linguistically used (‘phonological’) information should still be distinguished from gradiently varying (‘phonetic’) information in intonation. This is the central tenet of the Autosegmental-Metrical approach to intonation (Pierrehumbert 1980) which underpins most current research in intonation. Our data comprise a set of young, male speakers of SSBE for whom we have taken speech data from a series of realistic read dialogues, in which textual cues such as punctuation, context and stage-type directions prompted the speakers to produce tokens with an expression of more or less surprise. Linguistic function was varied by eliciting rises and falls in statements, questions and continuation contexts. At this stage we focus on data for pitch excursion, intensity and peak (or valley)-syllable alignment, investigating the relationships between categorical and gradient variation in these parameters as a function of their paralinguistic or linguistic functions. Initial findings indicate that an increase in pitch excursion plays an important role in marking surprise in H*L declarative tokens as distinct from emotionally neutral H*L declarative tokens, whereas for L*H question tokens, the difference in excursion is less pronounced.

This investigation forms part of a wider ESRC-funded project (RES-061-25-0347) which investigates the perceptual effects of certain acoustic properties of intonation patterns, as well as the brain systems that are involved in processing lower-level sound-based information, and higher-level, more abstract aspects of intonation.

Compensation in prosodic manipulations

Timothy Mills & Alice Turk
University of Edinburgh

This paper presents a compensation study looking at the production of emphatic accent. Studies of motor control reveal a pattern of goal-oriented compensation. In bite-block studies, rapid compensation for perturbation of jaw movement is exhibited in the movement of the lips to achieve closure (Folkins & Abbs 1975, Smith & Muse 1987) and of the tongue to achieve appropriate vowel formants (Lindblom, Lubker & Gay 1979, Fowler & Turvey 1981, Gay, Lindblom & Lubker 1981, Kelso & Tuller 1983, McFarland & Baum 1995, Baum & McFarland 1997, Baum 1999). When the movement of one articulator is constrained, the others compensate to achieve the functional target. The multiple joints involved in hand position (shoulder, elbow, and wrist) cooperate in a similar fashion (Flash & Hogan 1985, Flash 1987), suggesting that this compensatory behaviour is a general property of skilled action.

In the current study, we investigated the acoustic parameters used to express contrastive emphasis in English. We wished to see whether constraining one acoustic correlate of contrastive emphasis, f_0, would elicit a compensatory increase in the other correlates, duration and amplitude. We constrained f_0 in two separate manipulations: by eliciting sentences in whispered speech (removing f_0 information altogether), and by eliciting pitch accents in the context of sentence-final question intonation (leaving less f_0 range available with which to express contrastive emphasis). Six adult native speakers (4 female, 2 male) of Southern British English were recruited. All speakers produced high pitch accents in both the statement and question conditions, allowing meaningful comparison of relative f_0 rises. Speakers read items aloud, one at a time, as they were presented on a computer screen. Contrastive emphasis was indicated to participants with context sentences. For example, the word “surfing” in item (1) bears contrastive emphasis; in item (2), it does not.

(1) They’re not going swimming. They’re going surfing.
(2) They’re away today. They’re going surfing.

Ten contrastive/non-contrastive sets were devised, each using a different utterance-final trochaic-stress word as the target for contrastive emphasis. Items (3) and (4) illustrate the minimal segmental modifications used in the question manipulation.

(3) They’re not going swimming. Are they going surfing?
(4) They’re away today. Are they going surfing?

Each participant produced five repetitions of each item over the recording session. The whisper manipulation had no effect on duration or amplitude: they varied with contrastive emphasis by the same magnitude in whispered statements as in normally-spoken statements. The question manipulation showed a sympathetic effect. The f_0 correlate of contrastive accent was reduced in questions relative to statements (as expected). Rather than compensating for this loss, both the duration and the amplitude correlates were also reduced in questions. This pattern of sympathetic variation in physically independent parameters does not appear to have a parallel in the motor control literature, either for speech or for other skilled actions.

Four levels of tonal contrast in Peninsular Spanish boundary tones

Eva Estebas-Vilaplana
Universidad Nacional de Educación a Distancia

The main aim of this study is to provide evidence for the existence of four contrastive tonal levels in sentence final position (boundary tones) in Peninsular Spanish. Recent studies on Spanish intonation within the Sp_ToBI annotation system have incorporated two additional tones, a mid tone, M% (Beckman et al. 2002) and an extra high tone, HH% (Estebas-Vilaplana and Prieto 2008) to the original boundary tone inventory proposed in former investigations within the Autosegmental-Metrical framework (Sosa, 1999) which only included a low tone (L%) and a high tone (H%). This study presents the results of a production test in Peninsular Spanish which confirm the existence of the aforementioned four tonal categories in utterance final position.

A production test was designed including four kinds of sentences which were identical as far as the segmental and stress structures are concerned (i.e. *Manolo*), which were also produced with the same nuclear pitch accent (i.e. L+H* associated to the stressed syllable –*no*-) but which contrasted due to the different scaling of the boundary tones. Three Peninsular Spanish speakers from Madrid were asked to answer with the same word *Manolo* to different types of questions which prompted different intonational patterns, namely, a neutral declarative, an unfinished enumeration, a calling contour and an unexpected interrogative. Overall, 240 sentences were gathered. An acoustic analysis of the data was carried out by means of *Praat*. For each production, measurements of the F0 values were obtained at the following points: sentence initial and final positions, onset and offset of the accented syllable and highest F0 peak. A statistic analysis of the data was performed. The results showed no significant differences in the F0 values at any point in the four types of sentences and for all speakers except for final sentence position. These findings indicate the usage of four different tone levels in sentence final position in Peninsular Spanish to convey different meanings, that is, L% (declarative), M% (unfinished enumeration), H% (calling contour) and HH% (unexpected interrogative).

The results obtained in the production test corroborate the idea that Peninsular Spanish intonation cannot be accounted for by only two boundary tones (L% and H%), as it was proposed in former studies within the Autosegmental-Metrical model, but needs to incorporate two more tones to describe a four-level contrast observed in the data (L%, M%, H% and HH%). In the future, we expect to confirm this categorical distinction by means of perception tests.

A multimodal analysis of laryngeal constriction using videofluoroscopy and simultaneous laryngoscopy and laryngeal ultrasound

John H. Esling & Scott R. Moisik
University of Victoria

Laryngoscopy has allowed phoneticians to visualize the complex phonatory and articulatory actions of the larynx, but it provides only a limited impression of vertical movement. Changes in larynx height contribute to the configuration of laryngeal parameters influencing voice quality, pitch, and segmental articulations (Catford 1977; Laver 1980; Honda *et al.* 1999; Honda 2004; Esling & Harris 2005). Of particular interest in the present research, is how the larynx height parameter correlates with laryngeal constriction—the antero-posterior reduction of the epilaryngeal tube by means of aryepiglotto-epiglottal approximation. Laryngoscopy has been used to demonstrate how laryngeal height and constriction interact, but these demonstrations have only been qualitative (Esling 1996; Edmondson & Esling 2005). The present research seeks to address this deficiency by drawing on three different means of imaging the larynx: videofluoroscopy and, independently, laryngoscopy performed simultaneously with laryngeal ultrasound. These data are used to study the production of pharyngeal sounds to identify the quantitative relationships between larynx height and laryngeal constriction.

Videofluoroscopy of careful phonetic productions of [*ɑʡɑ*], [*ɑʢɑ*], and [*ɑʜɑ*] (with trilling in the fricatives) provides an opportunity to analyze and compare changes in the volumes of the pharynx and of the epilaryngeal tube. Changes in larynx height and in the vertical separation between the vocal folds and the aryepiglottic folds during stop and trill productions are also measured. Combined with similar productions filmed laryngoscopically, the data provide a better understanding of the articulations of the larynx that accomplish these gestures.

This research introduces a new laryngeal imaging procedure: simultaneous laryngoscopy and laryngeal ultrasound. Laryngeal ultrasound video data were obtained by placing the ultrasound probe on the subject's thyroid lamina, near the laryngeal prominence, while a standard laryngoscopic examination was being performed simultaneously using a flexible nasendoscope. As demonstrated in the medical literature, ultrasound is fully capable of generating an image of numerous laryngeal structures, which we will confirm. This research will illustrate how ultrasound can be used to track changes in laryngeal height using an automated image analysis algorithm based on the principles of optic flow analysis (implemented in MATLAB). This approach allows for large quantities of ultrasound video data to be processed without requiring manual intervention. Changes in laryngeal height relative to the ultrasound probe are quantified, as well as the velocity with which these changes occurred. The velocity data are then numerically integrated to determine the change in height of the larynx. Movements of structures in the laryngoscopic data are then compared with the laryngeal ultrasound data to provide a more robust picture of how the larynx is displaced during the production of sounds with laryngeal constriction.

This novel combination of three instrumental methods of laryngeal tracking together with the new data collected here will provide an enriched understanding of the vertical dimension of the laryngeal articulator in sounds that have long resisted image analysis.

Open-Mouthed or Stiff Upper Lip? Obtaining Data on Articulatory Settings in Bilingual Speakers

Sonja Schaeffler[1], James M. Scobbie[1], Ineke Mennen[2]
Speech Science Research Centre, Queen Margaret University, Edinburgh[1,] ESRC Centre for Research on Bilingualism in Theory & Practice, Bangor University[2]

As an abstraction lying behind the demands of producing phonologically distinct speech sounds, an articulatory setting can be described as a tendency for the vocal apparatus to keep returning to a language-specific habitual or neutral position in preparation for the next segmental gesture. Accordingly, an individual may have the tendency *"to keep the jaw in a relatively close position, or to set the lips in a habitually rounded position, or to have a rather whispery type of phonation"* (Laver, 1994: 115). Thus, articulatory settings are believed to shape broader and longer-term phonetic characteristics of a language as a part of the sound system.

There is plenty of anecdotal and impressionistic evidence supporting the concept of such settings, but hardly any empirical data that would quantify the nature or indeed prove the existence of articulatory settings.

In this paper we will discuss results of a methodological study we have undertaken to develop capacity for analysing cross-linguetic differences in articulatory settings systematically, in a large number of speakers and for a broad range of languages. We will present articulatory data from eight German-English bilingual speakers. For all eight speakers Ultrasound data was obtained to determine the overall shape of the mid-sagittal section of the tongue surface contour, together with simultaneous acoustic recordings. For four speakers we also gathered VICON data to obtain three-dimensional positions of the articulators and to determine lip spreading, lip protrusion and chin lowering (which we assume reflects jaw opening).

The two languages were tested in separate language blocks. Speakers were asked to read sentences out loud and were also administered a Picture Naming Task which elicited word pairs that are phonemically similar and with that 'comparable' across the two languages, German and English (such as Fisch/fish, Haus/house, etc). Recordings started 5 seconds before prompts appeared on the screen. During this delay participants were listening to a variety of naturalistic pre-recorded instructions presented via headphones (e.g. "And what's the word for the next picture?"). That way, preparation for speech was captured, and elicited in an ecologically valid set-up that resembles more closely the 'natural' switching between speaking and listening.

We will present articulatory measurements from both periods of acoustically realised speech and periods of preparation for speech, and test whether speakers keep their two languages apart by inducing measurable changes to their articulatory settings. We will discuss how the results impact on future directions of research on articulatory settings, and more generally how the design of experiments has to be modified to facilitate ecologically valid measurements of speech articulation.

Variable tongue configuration in Scottish English postvocalic-/r/.

Eleanor Lawson[1], Jim Scobbie[1] & Jane Stuart-Smith[2]
Queen Margaret University[1], University of Glasgow[2]

From an auditory perspective, a continuum of postvocalic /r/ variants can be heard in the Scottish Central Belt (SCB); from strong variants such as “retroflex” approximants, to weak variants where an audible /r/ articulation seems almost absent, Speitel and Johnston (1985), Romaine (1978), Stuart-Smith (1999), (2003), (2007). Variants at the weak and strong ends of the continuum are associated with lower and higher socioeconomic speaker groups respectively. Males and females also use different proportions of strong and weak variants, Romaine (1978); Stuart-Smith (2007). Romaine (1978) was perhaps the first to point out that the increased sociophonetic complexity of /r/ variation in eastern SCB speech pointed towards sound change in progress.

Ultrasound tongue imaging (UTI) allows us to look beyond the auditory level in order to gain a better understanding of the source of much of the variation we hear. A previous UTI study, Scobbie, Stuart-Smith and Lawson (2008), considered the timing of lingual gestures and its contribution to the perception of a weakened postvocalic /r/. It showed that raising of the tongue tip was still present in weakened /r/, but the tip-raising maximum was delayed to a point beyond the offset of voicing.

The present study focuses on differences of tongue *configuration* in the post-vocalic /r/ variants of our socially-stratified UTI corpus of young eastern SCB speech.
There are three main findings from this analysis:

1. We show that many of the variants that would traditionally be transcribed as “retroflex” do not involve a retroflex lingual configuration, or even tongue-tip raising.

2. We highlight deficits in the IPA system of auditory transcription for postvocalic /r/.

3. We show that tongue configuration for postvocalic /r/ is socially stratified in our corpus with speakers belonging to different socioeconomic cohorts and different gender groups using markedly different tongue configurations in the production of post-vocalic /r/.

Rediscovering the lost X-ray data of Tsutomu Chiba (1883-1959)

Michael Ashby and Kayoko Yanagisawa
UCL

This paper describes the discovery and identification of an undocumented collection of 89 glass lantern slides (dated roughly to the first half of the twentieth century), which contain superb midsagittal outlines of the vocal tract made from X-ray tracings. Fifty-seven of the slides, found in London, show target positions which constitute essentially a complete Japanese syllabary; the remaining 32 are devoted to English (RP) sounds. Though apparently unpublished and unknown, the images provide visual catalogues arguably more comprehensive and detailed than any others available for these languages, even today.

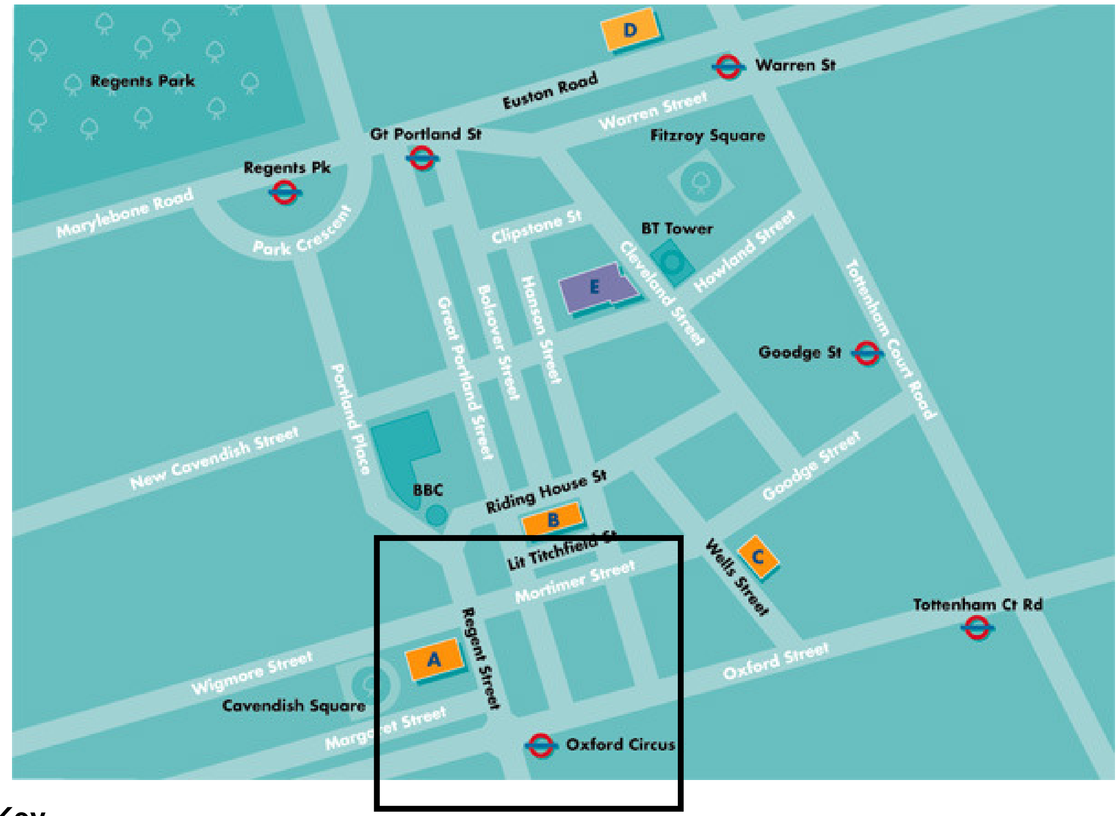
The figures show specific resemblances to some included in Chiba and Kajiyama (1942), especially in the detailed drawing of the vertebral column, epiglottis and hyoid bone, and a conjectured attribution to Chiba was finally confirmed when contact prints evidently prepared from the same negatives were found among Chiba's effects in Tokyo. Examination of Chiba's publications (rare even in Japan) suggests that the slides are not simply a record of illustrations in any publication, but rather that the slides represent the primary research data itself, and probably date from the period 1931-1934.

Though little known in the West, Chiba (1883-1959) was one of the most significant of all Japanese phoneticians, founding a superbly equipped phonetics laboratory in 1929 at the Tokyo School of Foreign Languages – "probably the best…in the world" according to Palmer (1936) – and with his co-worker Masato Kajiyama (1909-1995) pioneering the acoustic theory of speech production about a decade ahead of corresponding developments in the West (Chiba and Kajiyama, 1942). It has hitherto been believed that all of Chiba's research data was destroyed in the fire-bombing of Tokyo in spring 1945 (Maekawa, 2002).

Chiba studied in London (1913-1915) and maintained a connection with the IPA, but it remains to be explained how the collection of slides came to be in England. Various indications point to the possibility that they were prepared and shipped to London for a planned presentation at the second ICPhS, which took place in 1935, though in the event Chiba did not attend that conference (Palmer, 1936).

Index of presenters

UNIVERSITY OF WESTMINSTER LOCATION MAP



Key

A 309 Regent Street - School of Social Sciences, Humanities and Languages, Gym and Sports Hall, The Deep End