Looking variation and change in the mouth: developing the sociolinguistic potential of Ultrasound Tongue Imaging

Research Report for ESRC Project RES-000-22-2032

James M Scobbie, Jane Stuart-Smith and Eleanor Lawson

June 2008

Background

The central goal of this project is to meet a pressing need: to enable the investigation of how speakers from anywhere on a socio-dialectal spectrum physically articulate speech.

On the one hand, we need to augment sociolinguistic acoustic data with articulatory data (Kerswill and Wright 1990). This is an end in itself, but also lets us address the complex relationship that exists between the sounds of speech and the vocal tract configurations that generate them in a wider context than normal. Normally such issues are the concern of a tradition in instrumental phonetics which concentrates on standard dialect speakers in a formal laboratory setting. Research may focus on acoustic invariance (e.g. Guenther, et al. 1999, Mielke, Baker and Archangeli in press), functional articulatory tendencies (e.g. Sproat and Fujimura 1993) and/or the perception of lexical (but not social) meaning (e.g. Silverman 1995, Gick et al. 2006). Social variation cannot be understood fully, we contend, unless it is approached from all these perspectives.

On the other hand, the narrowing of the empirical base of experimental phonetics means that the social function of phonetics is absent as a topic of study (Foulkes and Docherty 2006), along with discourse and interpersonal functions (Local 2007). This means that the fine-grained systematic differences that exist between dialects and sociolects hang out of reach, and are not used as a typological tool by linguistic theorists who are, nevertheless, interested in gross cross-linguistic variation. Only through the analysis of subtle variation can progress be made in understanding the formation of linguistic categories and linguistic systems (e.g. Scobbie 2006), and dialectal/sociolectal variation provides the ideal type of phenomena.

The challenges lie in blending best practice in sociolinguistics with innovation in instrumental phonetics in such a way that all interested parties can benefit, across a broad spectrum of theoretical and empirical interests.

Until recently, the difficulty of collecting articulatory data from naïve vernacular speakers has largely been an insuperable impediment to research in both sociolinguistics and phonetics. A generation has passed since the pioneering work (in ESRC-funded projects C00232227 & R000231056) by Kerswill, Nolan and Wright from 1985-1991. Although they convincingly showed the value of vernacular articulatory research, specifically through the use of electropalatographic (EPG) data on tongue-palate contact (Kerswill 1985; Kerswill and Wright 1990; Wright and Kerswill, 1989), this particular technique has not been adopted in sociolinguistics (despite the continued

potential: cf. Scobbie, Pouplier and Wrench, 2007), and social aspects of phonetic research are largely confined to gender and age differences.

EPG is relatively expensive on a per-subject basis, since an individually-made palate is required, and this is probably the main impediment to its adoption — the data is easy to interpret and quantify. Ultrasound Tongue Imaging (Stone 1997), on the other hand, has many properties required by a fieldwork tool (Gick 2002) and might be prima facie suitable for sociolinguistic research.

The specific phenomenon addressed here is the behaviour of Scottish English /r/, which previous work (e.g. Stuart-Smith 2007, Scobbie, Stuart-Smith and Sebregts 2006) suggested would provide theoretically important data, if socially-interpretable and naturalistic articulatory data (both spontaneous speech and experimentally controlled) could be gathered using Ultrasound Tongue Imaging.

Objectives

All were met.

Objective 1. *A methodology for articulatory sociophonetic research using non-intrusive direct measurement of speech articulation.*

We have facilitated future research using vernacular and socially-stratified articulatory data. Specifically, we employed Ultrasound Tongue Imaging (UTI) with a headset to stabilise the probe, and with provably-synchronised acoustics. We have developed methodological protocols in a number of areas such as: research ethics, recording procedures to assist in the elicitation of vernacular speech, data collection hardware and software (including use of the stabilising headset for ultrasound), analysis software and techniques (Methods and Results). We conclude that the vernacular may be elicited even with the articulatory equipment in place (Output #1 and Results). Our work on **Objective 4** demonstrates the value of such data (cf. Results, Impacts).

A resource containing protocols, techniques and examples is available to research groups. Suitable examples are being added for the benefit of research subjects and the increasing number of interested researchers at the website
http://www.qmu.ac.uk/ssrc/ultra/protocols/ and more generally
http://www.qmu.ac.uk/ssrc/ultra
(See Annex 2 for a reproduction of the index page).

These protocols will form part of our contribution to two chapters to a textbook on best practice in sociolinguistics (cf. Dissemination).

Manuals for the UTI hardware, software, and headset are available on request from
http://www.articulateinstruments.com

Objective 2. *The first socially-stratified articulatory/acoustic dataset*

We collected the East Central Belt 2008 UTI corpus "ECB08" as Phase 3 of our recording schedule. See Method, Results and Annex 1. The *core* structure of ECB08 (Table 1) can be augmented (see Annex 1 and Future Research Priorities).

Table 1 Social stratification of ECB08 showing numbers of subjects

age 12-14	Working Class (West Lothian)	Middle Class (Edinburgh)
male	4	3
female	4	4

Earlier recordings constitute an extra corpus, the West Lothian 2007 UTI corpus "WL07". For details of Phases 1 and 2, see Annex 1 and Method. WL07 comprises field and laboratory recordings of 14 working-class boys (12 recorded with UTI) from the same school (and year-group) as the 8 WC speakers in the ECB corpus. ECB08 is of better technical quality than WL07.

We thus have a mixture of different data types and recording locations as well as stratification for gender and social class. This is the first articulatory dataset so structured in the UK, and we believe internationally. ECB08 has equal focus on spontaneous speech (articulatory data from half of the participants) and experimental materials. WL07 was, as planned, subject to a quantitative and qualitative auditory-acoustic analysis (Output #1).

Objective 3. *Sociolinguistic and phonetic analysis of Scottish derhoticisation using acoustic and articulatory data*

We have been able to confirm our initial hypothesis about the complex relationship between articulation and acoustics in WC derhotic speech (Output #1, Results). Speakers may indeed have strong but covert anterior gestures in utterance final position in derhotic speech which have negligible effects on the acoustic rhoticity of the output, resulting in near homophony between similar words which differ in the presence/absence of /r/. See Results and Output #1 for more details.

As envisaged at the outset, the large amount of data collected for the corpus, the time-consuming nature of UTI analysis, and the short timescale of the project means that analyses are preliminary. See also the sections on Difficulties in the End of Award Report Form, and Future Research Priorities. Due to the change of the fieldwork location from Edinburgh to West Lothian, the QMU campus move, and the decision to record the corpus in the laboratory, data collection was delayed. Another factor making analysis of derhoticisation difficult is that there were simply fewer derhoticising speakers than expected among the WL07 children.

Objective 4. *Theoretical discussion and development of future research goals*

The basic position we are working on is as follows (and see Output #2, Results, and Future Research Priorities).

People are socially-situated speaker-hearers. To understand a person as a speaker, we need articulatory data. To understand them as a listener, we need perceptual data. To understand them as a integrated whole, we need to see them as an "agent" who is aware of the effect of their own acoustic output on other listeners and who use their system accordingly, and as someone who actively monitors the systems of others. Acoustic analysis only allows us to study speakers' output, but not the articulatory strategies underlying that output; and as our data show (Results), these two levels may not always

align as "neatly" as might be assumed. As we have shown here, articulatory data has the potential to reveal more "internal" aspects of the output than acoustics. Acoustic output *is intended to be heard*, but our articulatory data shows that "perceptual recoverability" of lexical and social meaning is more complex than is normally thought.

All phonetic systems and the contexts in which they are studied by linguists are socially-embedded. A homogenous sample of middle-class, highly literate, standard dialect speaking colleagues is just as much a social sample as a group of vernacular school-children. A major problem for advancing research in our fields (independently and in collaborative work) is that different groups of researchers working on different theoretical questions tend to use different methodologies on different types of linguistic systems. We introduce confounds into our research if studies of social variation and vernacular systems favour corpus studies based on spontaneous speech samples, while controlled experiments relying largely on participants from highly educated middle class speakers remain the preserve of experimental phoneticians. When methodological practices and research questions become firmly linked, the opportunities for new understanding and the cross-fertilisation of ideas are harder to achieve.

A counter-view to ours (Thomas, 2002: 168) is that only *acoustic* instrumental analysis is relevant for sociolinguistics. This is a listener-oriented theory of variation, in which it is only what people hear that matters. It is ironic that this view explicitly relates to variation in (American English) /r/. Our preliminary results on Scottish /r/ strongly argue that articulatory, acoustic and perceptual evidence are all essential independent elements. This may be particularly for speakers accommodating to a new pattern, or it may be inherent to all variation: such questions remains to be investigated. What is clear is that people are both speakers and listeners, and that they are required to systematise the variation which comes to them as input and then produce their own output (which will, moreover, be systemically variable in its own right).

Methods

One major aim of the study was to find out the potential psychosocial impact of ultrasound recording, particularly when it involved the wearing of a stabilising headset (Figure 1), the presence of several pieces of hardware and 2-3 researchers setting up and monitoring the recording. In what ways would this overt observation make subjects behave differently than they would in the presence of a microphone for audio-only recording?

In answering this question, we needed to develop means to minimise and/or track such differences (**Objective 1**), which have led to our suggested protocols for collecting sociolinguistic and/or vernacular articulatory data (Annex 2).

To test the effect of Ultrasound Tongue Imaging on speakers, we originally planned to record 12 females with either headset (Figure 1) or hand-held probe (Figure 2) and ascertain the differences these conditions caused in levels of vernacular speech variables relative to audio-only recordings. From pilot work and scoping in schools and other venues, we decided to use only the headset condition, which provides more reliable data, and because, if this condition were to produce no particular change in speaker behaviour, then hand-held UTI would be likely to be similarly neutral.

Figure 1 Headset to hold probe steady and allow natural head movement during speech. Despite resembling a constraining medical device, the headset feels like a sports helmet and was generally regarded by subjects as amusing and "cool".

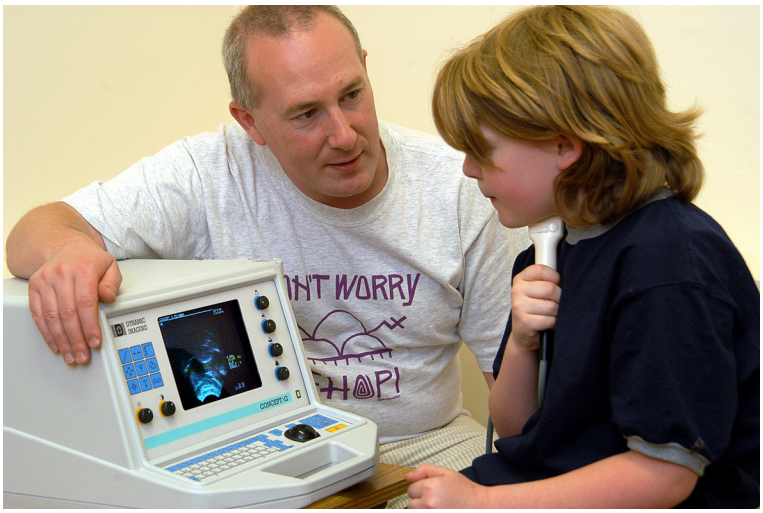


Figure 2. Hand-held probe in use with 8 year old child

As planned, participants were recorded in friendship pairs. They were left unattended with instructions just to talk to each other (see the protocols on the website, Annex 2). for the spontaneous speech (SS) sample. If they were being recorded with ultrasound, both participants wore full headset and probe, though only one was recorded (Figure 3).[1]

[1] The figure shows two adults because we did not seek permission to take photographs of the child participants. All photographs shown in this report and on the website have permissions.

Figure 3. Two adults wearing Articulate Instruments headsets during spontaneous discourse. A tabletop microphone is shown, but lavalier microphone were used.

We decided to record young males for Phases 1 & 2 (Annex 1) because scoping in Edinburgh indicated that more males than females might derhoticise. Having identified suitable schools in Edinburgh using demographic data, we were denied access (see End of Award Report Form), requiring us to use a different local authority. West Lothian was an interesting and suitable location, and following an assessment of local demographics (Scottish Index of Multiple Deprivation 2006) access was granted by the local authority and head teacher to the WC school of choice, selected on a number of indicators.

Throughout the project, there was a tension between the informality of the recording location and protocol (judged by us) and the quality and usability of the recordings. Data collection methodology underwent several modifications in order to optimise collection of good-quality ultrasound and audio data (Annexes 1 & 2). Phase 2b involved UTI data collection in the field, requiring our laboratory equipment to be transported to the school.

One researcher (EL) recorded 14 informants with audio. The next week, three researchers (EL, JS, SC) re-recorded 10 of these original participants under UTI recording conditions, or, for the controls, with audio only by EL alone without any UTI equipment being present. This methodology allowed us to compare behaviour in the UTI condition to the participant's own audio-only recording, while controlling for the effect of re-recording.

Output #1 provides a detailed analysis of differences between Phases 1 & 2. We summarise here, and add comments on Phase 3. A comparison was made of easily-transcribed socially-salient variables (TH-fronting, T-glottalling, and L-vocalisation) in word list speech in Phase 1 and either Phase 2a (audio only, control condition) or Phase 2b (with ultrasound). From the spontaneous speech sample, the amounts of /r/ derhoticisation were analysed (excluding the first five minutes of the recording). There is no clear pattern of variation between the control group and the group recorded under UTI conditions. In particular there is no overall average decrease in vernacular variables in the UTI condition. Individual subjects vary.

However, we informally observed that some of the Phase 2a participants seemed more "at ease" than those in Phase 2b. Our impression was that this might be detected by quantifying the amount of talking, their "loquacity". Therefore, in addition to the planned comparison of variables, words per minute and total output were analysed for the *first* five minutes of interaction (Lawson in preparation). *All* speakers spoke less during Phase 2, but the drop in loquacity was greater for control speakers being recorded with audio only. Comments made during the spontaneous speech recordings by participants regarding the difficulty of finding things to say in the audio recording scenario suggested to the researchers that the introduction of UTI itself might have given the informants a new stimulus for speech.

Overall, the participants' reactions to the ultrasound recording scenario were idiosyncratic, with one pair talking freely and excitedly throughout, as they had done during the audio–only recording, while another pair spoke far less. The interpersonal dynamic of the conversational partners seemed to have a bigger effect on the individuals' behaviour during the UTI recording scenario than the scenario itself. That is to say, some conversational pairs adapted more quickly to the unusual recording scenario than others. Those who wanted to chat, were happy to chat (regardless of the setting), while others whose relationships were more strained seemed to find it hard to find a topic of conversation.

It must be added that the UTI recording scenario was not just "formal and scientific": informants who volunteered for the experiment were allowed to chat with a friend and drink juice at a time when they would normally be required to be in class. They were also aware that they were objects of interest to researchers and received individual attention from researchers during the recording process. The unusual status of the (self-selected) research participants among their peers was not wholly lost on them, and the kudos was clearly greater for those who had a day out at the laboratories.

An obvious problem associated with carrying out both audio and ultrasound recording in the field will always be the quality of the recordings (both audio and ultrasound). For example, having equipment such as a PC and an ultrasound machine near the microphone leads to extraneous noise on the audio recordings. Even though a comparatively quiet room in the school was reserved, it had bare flat walls and lacked acoustic damping, characteristics which augmented the continuous computer noise and other extraneous noise with reverberation and echo. Extraneous noise was caused by events such as heavy rain drumming on skylights during a thunderstorm, school bells, and interruptions.

We evaluated these factors, and decided, in light of our specific interest in derhoticisation in utterance final position (i.e. in the transition from speech to silence) that we should record our final corpus (in Phase 3) in a laboratory setting. Recording studios were, however, not available at the time we needed them for the corpus, due to QMU's campus move (see the End of Award Report Form); so the focus of the project turned for a while onto analysis and dissemination from WL07. As a final check, Phase 3 was piloted in the laboratory, in Phase 2c. Impressionistically, we did not detect any difference in the behaviour of these four subjects from their previous participation in Phase 2a & 2b, though formal analysis has yet to be undertaken.

A further change of plan was required when our chosen West Lothian MC school found they were unable to participate. Rather than approach a less clearly demarcated West Lothian MC school, we decided a school within Edinburgh City would provide a clearer MC sample: the school (a fee-paying private sector school) was very co-operative and happy to participate in the research.

To record the ECB08 corpus, participants came to QMU (in Musselburgh) by taxi in same-sex groups of four. The WC speakers were from the same school and year-group as the previous West Lothian participants (now in 2nd year). All were naïve as to the nature of the linguistic research topic and had not been previously recorded (Annex 1).

Results

The main results are as follows

1. UTI is a feasible sociolinguistic tool and protocols have been developed.
See Lawson, Stuart-Smith and Scobbie (in press) (Output #1)
The use of the ultrasound tongue imaging equipment, including a stabilising headset, did not cause a large or consistent style shift in the speech, or at least no greater than that which occurred through the re-recording of a control group with a standard microphone and audio-recording equipment.

2. Derhoticisation data collected in the laboratory is representative of natural derhoticisation behaviour in spontaneous speech.
We have observed derhoticisation with covert tongue movements in a range of different speech styles and sociolects, and we are confident it is not an artefact of the UTI method (Figures 4-6). Acoustic analysis confirms a lack of the acoustic features most commonly associated with a canonical Scottish rhotic coda consonant, namely raising of F2 and steep lowering of F3 (compare spectrograms 1 and 2 from rhotic tokens with spectrograms 3 and 4 from derhoticised tokens in Figure 4).

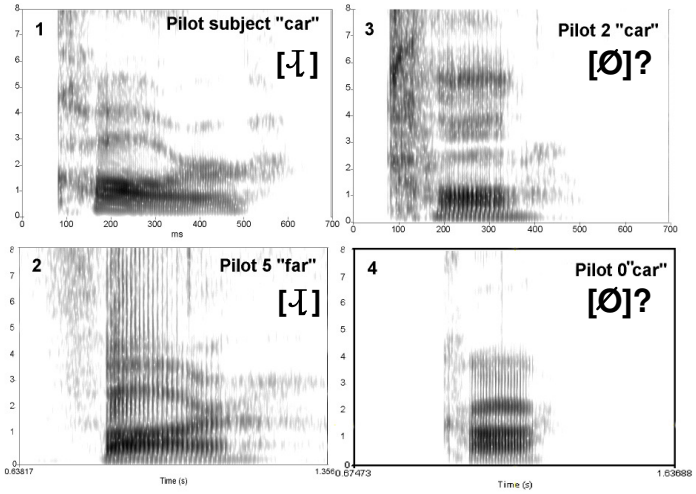


Figure 4 Spectrograms showing four separate speakers (pilot stage) producing words with coda /r/. Left hand images (1,2) are of rhotic pronunciations; right-hand images (3,4) show derhoticisation.

Although spectrograms 3 and 4 in figure 4 show flat formants, listeners may report hearing them as (weakly) rhotic, just as they do for diphthongs ending in a schwa or

pharyngeal offglide. An analysis of UTI images of these recordings revealed a delayed tongue-tip raising gesture, examples of which are shown in Figures 5, 6.

.

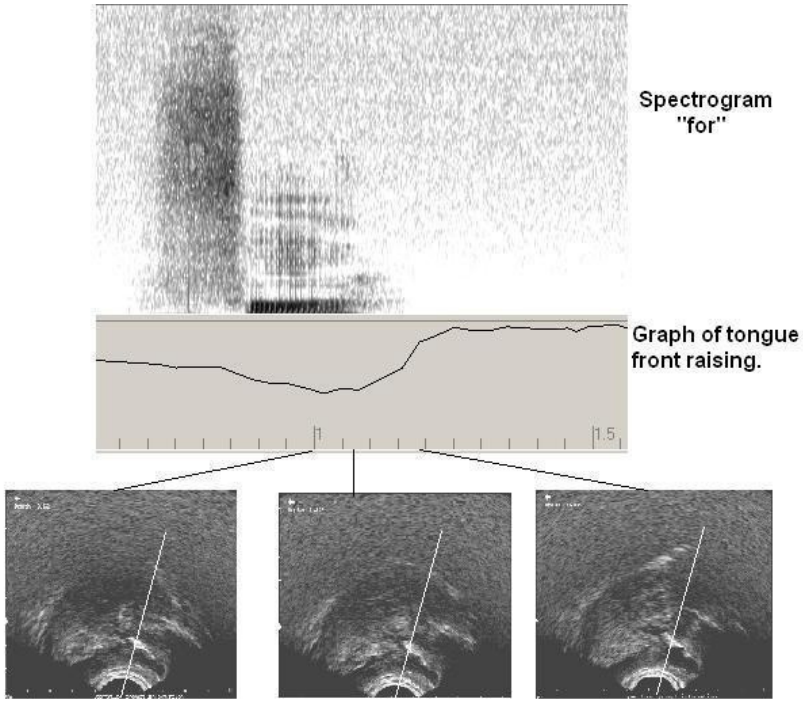


Figure 5. Bottom panel: three UTI frames of Pilot 2 (24 year old male from West Lothian) saying *for* (heard variably as [fɔ] or [fɔɹ]). Tip raising along the white trajectory line is graphed in the middle panel, time aligned (in seconds) with the spectrogram in the top panel. Note the breathy source at the end of phonation between the second and third UTI frames.

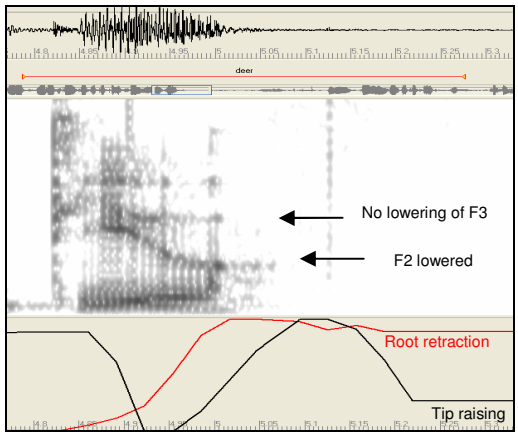


Figure 6: waveform of derhoticised *deer* [di□] from WC male speaker from the WCB08 corpus, drawn from the spontaneous speech sample. The bottom panel shows two time-aligned analysis values of tongue movement.

With UTI, singleton /r/ could be seen as a strong anterior curving post-alveolar gesture. Such tongue raising gestures can be so delayed that they begin during the last few cycles of phonation or even after it has ceased (see Figure 5, Figure 6). The auditory impression is that /r/ is phonetically weak or absent: the cavity change might be inaudible or there may be glottal source energy from a breathy whisper in which case

final /r/ might be said to be a low intensity voiceless fricative. Our results may explain problems encountered by trained phoneticians when attempting to agree on the rhotic quality of variants produced by Glaswegian adolescents (Stuart-Smith 2007).

3. Pre-pausal utterance final position favours derhoticisation
This context has been noted as particularly conducive to derhoticisation in Edinburgh WC speech by Romaine (1979:45) and Speitel and Johnson (1983: 28). Figure 7 shows the pattern is also true of WL07.

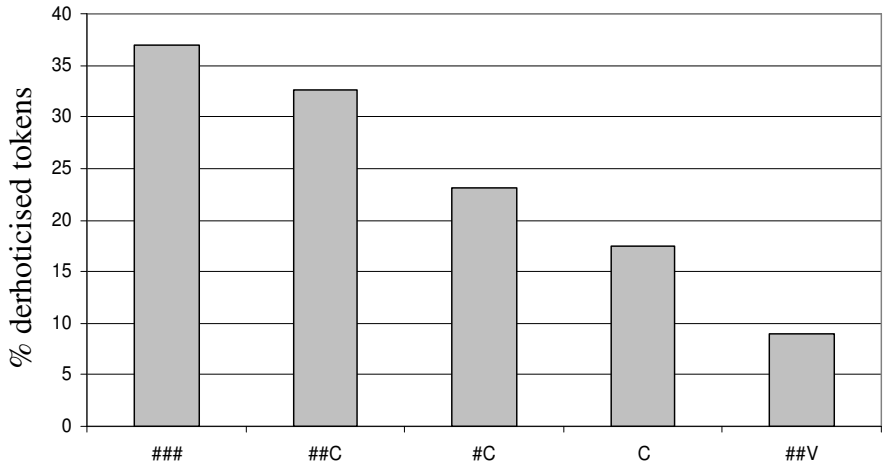


Figure 7 The effect of phonological environment on derhoticisation in spontaneous speech. The contexts are prepausal utterance- final (###), word-final before a following consonant (##C), word-internal syllable-final before a following consonant (#C), in a word-final cluster (C) and word-final prevocalic (##V). WL07 corpus, n= 2567.

As noted, tongue-root retraction occurs relatively earlier than tongue-tip raising (see Result 2, Figure 4, Figure 5, Figure 6). The temporal separation of these gestures in our UTI data appears to be an extreme example of gestural dissociation (cf. Sproat and Fujimura 1993). Perceived derhoticisation occurs most frequently in utterance-final position which might be explicable if this is where dissociation is strongest, though pre-pausal utterance final position (e.g. Oller, 1973) has been somewhat neglected in recent studies of such prosodic effects in favour of phrase-medial positions (cf. the criticisms in Gordeeva and Scobbie 2007). Even if a general process like gestural delay or clock slowing (Byrd and Saltzman 2003) is the mechanism behind the early stages of vernacular sound change in vernacular Scottish English /r/, it is crucial to understand that because the behaviour is socially stratified, it is demonstrably not automatic.

4. Derhoticisation is most common in weak syllables in West Lothian WC speech.
Derhoticisation in West Lothian was far more common in weak syllables (Figure 8). Weak syllable derhoticisation was also reported in more widely derhoticised Glasgow speech (Stuart-Smith 2003, 2007). In weak syllables, we think there is less likely to be a covert tip gesture, but since UTI with a sampling rate of 30Hz is less suitable for the analysis of small or rapid movements of short duration, we have concentrated in the ECB08 wordlist on /r/ in fully stressed syllables. Derhoticisation is likely to be due to more than one articulatory mechanism, just as rhoticity is coarticulated.

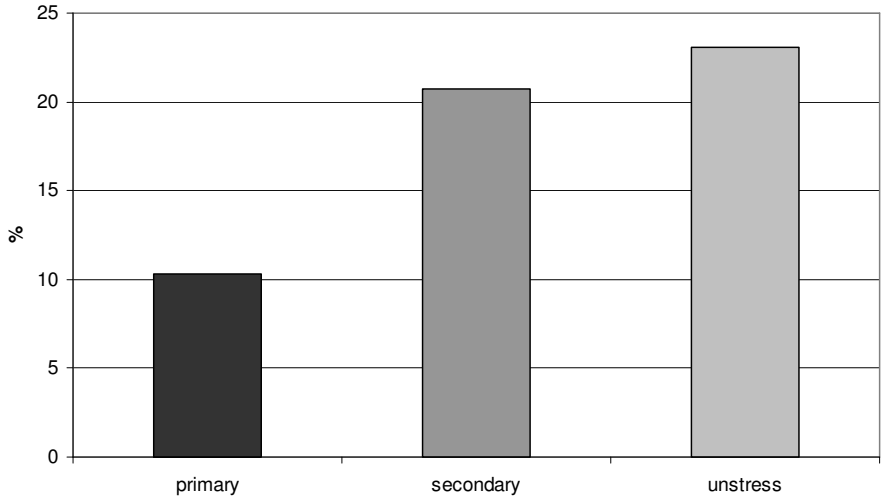


Figure 8 The effect of syllable stress on derhoticisation in spontaneous speech. WL07 corpus, n= 3217.

5. Derhoticisation in low vowels may lead to near merger and/or new vowel phonemes among low back monophthongs

In rimes /ar, □r/, WC derhotic productions are mostly monophthongal. Our experimental study focussed on these low back vowels, to see how words like *hut*, *hurt* and *heart* are distinguished utterance finally. In derhoticising speakers, *hut* and *hurt* are short, *heart* is longer (vs. *hat* and *bra* which are front with a quite different vowel quality). Speculatively, this may lead to an /a/ vs. /□/ split with interference between /□/ and /□/, and/or merger of /□r/ and /□/ in some closed syllables.

In our MC Edinburgh speakers, who are all rhotic, /□r/ is rare, so *verb*, *berth* etc. have largely merged with *fir*, *fur*, *birth, hurt*, with a strong rhotic vowel quality confirming Aitken (1979: 111). This rhotic monophthong seems to be salient sociolinguistically, and we can differentiate WC vs. MC *verb*, *hurt* ([v□ˑb], [h□□, h□ˑt□] vs. [v□b] [h□t]). Perhaps the latter have /□/ phonologically: but in any case, phonological compression in the number of possible rimes does not seem to be more advanced among derhoticised speakers.

6. Strong breaking was observed after mid and high vowels

An epenthetic vowel between high and high-mid pre-rhotic vowels and coda /r/ has a long history in Scottish English (Aitken 1979: 103-4). Such breaking was particularly prominent in utterance-final position (cf. Romaine 1979: 145). Our articulatory data confirms Lindau's (1978, 1985) suggestion that breaking is a result of an early onset of the tongue root retraction gesture that is present in postvocalic /r/ (Figure 6), but breaking is not an automatic transition between a vowel and a liquid (Gick and Wilson 2001; Gick et al 2002). It is a WC sociolinguistic target in its own right, which appears to be taking on an increasingly important role for some speaker-hearers as a cue to the contrast between those words which do and do not contain coda /r/ in English. In this way, derhoticisation may (gradiently and/or categorically) be turning into non-rhoticity via the creation of diphthongs.

7. Different tongue shapes are used to articulate /r/, with contextual and speaker variation.

Quantitative and qualitative analysis of these patterns is ongoing, but the idea that a rhotic Scottish /r/ is a "retroflex" approximant can be strongly counterexemplified: the situation is far more complex.

Activities and Outputs

Outputs #1[2] and #2 arose from conference presentations. On the ESRC Society Today site eight additional unpublished conference papers appear: in total we presented to conferences in Scotland (1), England (3), Northern Ireland (1), Germany (1), USA (3) and New Zealand (1).

The Outputs were peer-reviewed. Output #1 was invited on the basis of the oral presentation at NWAV on which it was based and Output #2 arose from an invited presentation at an ICPhS symposium.

The End of Award Report Form in the Dissemination section lists planned outputs. Three are highlighted here because they will directly make use of data and concepts from this project, namely the Yaeger-Dror et al. chapters and Scobbie, Stuart-Smith and Sebregts (in preparation). The former were invited by the editors. The latter is based on an invited keynote talk (at the OnLI conference) and a Labphon poster from 2006. An invited oral presentation Scobbie (in preparation), and the published presentation Scobbie, Wrench and van der Linden (in preparation) will report this project.

Scobbie has organised a symposium at IASCL on instrumental methods in child language acquision, and will present results from the projects.

This project has helped in the preparation of Foulkes, Scobbie and Watt (under review).

Impacts

Dissemination and networking have been well received in sociolinguistic and UTI research communities. Our commitment to such dissemination can be seen in the range of audiences with which we have attempted to engage.

Following a research visit by Malcah Yaeger-Dror (Arizona) during a longer sabbatical at York University, our UTI protocols are being incorporated into a new text on best practice in sociolinguistics.

Future Research Priorities

A number of new research questions arise on the phonetics of Scottish /r/ and Scottish derhoticisation, both among fully rhotic and derhoticising speakers, such as variation in tongue kinematics and the extent of derhoticisation in a wide range of prosodic and phonotactic contexts.

[2] Publication due 2008.

In terms of **variation and change**, we found derhoticisation to be weaker in West Lothian (in the east of the populous Central Belt of Scotland) than is reported for the western part: WL07 indicated an average 20% derhoticised tokens in spontaneous speech vs. over 70% in Stuart-Smith et al.'s (2007) Glasgow corpus. Indeed, aspects of derhoticisation in Glasgow may be almost complete for certain vowel contexts for certain speakers. Perhaps the phenomenon is more advanced in the west, both in terms of the number of speakers, and in terms of the articulatory basis of the impression, than in West Lothian. It may be also that WC urban Edinburgh speakers are more derhotic than WC West Lothian speakers. UTI studies of **derhoticisation** in vernacular urban Edinburgh speech and West Central Belt varieties emerge as priorities. Furthermore, data on Anglo-influenced MC non-rhotic speech would strengthen our understanding of the differences between categorical non-rhoticity and derhoticisation. Studies of Scottish rhotic speakers would make a very usefully addition to the literature on **rhoticity** which is so dominated by American English. But the priority ought to be the patterns and detail of derhoticisation. Longer-term, our corpus will be important for **real-time study**. ECB08 corpus will be a baseline for derhoticisation among vernacular speakers, precisely because we know already that, in comparison to Glasgow, say, the phenomenon is weaker and confined to fewer environments, but is likely to increase in generality.

In terms of **phonetic theory**, the utterance-final context has emerged as being of particular interest, a factor which fits in with other articulatory and acoustic work by Scobbie and colleagues (Pouplier, Gordeeva) on /l/ vocalisation, pre-aspiration of fricatives and ejective realisations of stops, suggesting that a generalised study into gestural dissociation of supralaryngeal and laryngeal gestures in different segment and cluster types pre-pausally would be a very valuable step.

We need to develop new methods to **quantify ultrasound patterns**, and this is clearly a priority, now that we have stabilised experimental data. We have made an excellent start in this area, using fan-grid to measure extent and relative timing of gestures (Figure 9). Formal analysis of our experimental materials using quantitative methods and via impressionistic analysis of tongue shape kinematics is underway already.

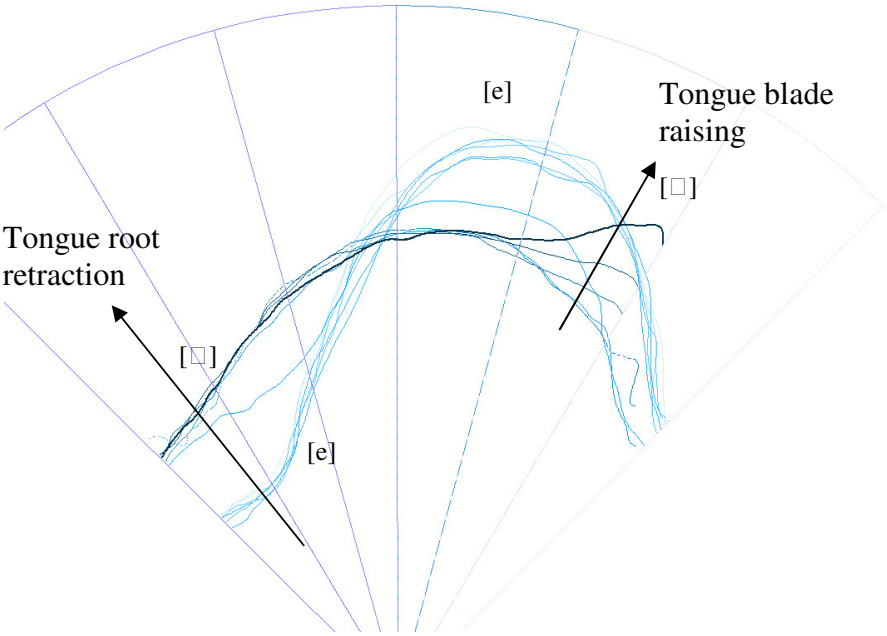


Figure 9 Waterfall pattern showing temporal change in tongue shapes during the rime of the word *hair*, from early [e] frames to the final [□] one. Tip raising occurs temporally after root retraction, enabling it to be nearly inaudible. The reference arrows show the dimensions in which time-varying constrictions can be calculated for quantitative analysis

It seems also essential to study **perception**, to add to our theoretical understanding of the relationship between speaker-hearer activities in the individual.

In terms of **technological development**, the most relevant task is to trial truly portable UTI equipment to facilitate fieldwork, particularly with the aim of gathering large samples in the field, e.g. at the Glasgow Science Centre.

In terms of the study of **spontaneous discourse**, articulatory data provides information on *inaudible* movements in the listener which sometimes indicate an unrealised change of turn. This is a topic ripe for future collaboration with discourse and talk-in-action specialists. We are taking this topic forward as part of the EPSRC "Edinburgh Speech Production Facility", where Electromagnetic Articulography will be used with *two* speakers in semi-spontaneous discourse. In general, we need to examine the UTI data from spontaneous discourse to search for new phenomena, as well as exploiting UTI as a tool capable of eliciting relatively relaxed and vernacular speech.

Word count 4,976

Annex 1
Phases of Data Collection and Corpora

Phase 1&2 = West Lothian 2007 UTI corpus "WL07"
Phase 3 = East Central Belt 2008 UTI corpus "ECB08" (the project deliverable)

Total hours of ultrasound + audio data:[3]

Spontaneous speech	**2h. 14mins.**
Word list	**2h. 10mins.**

Total hours of audio data:

Spontaneous speech	**15h 24mins.**
Word list	**2h 23mins.**

The following tables give information about subjects, subject groups, and the nature of data collected. In ECB08, one MC male participant was absent from school on the date of recording, so *n*=15 for the corpus. UTI discourse data (SS) was collected for 8 participants as planned: one non-recorded interlocutor doubled up due to the absent participant. In addition, the MC speaker group includes an extra post-hoc convenience sample of slightly younger children (n=4) known personally to the P.I. from a socially comparable state-sector (primary) school (from P6, two year groups lower than the core ECB08 MC sample of S1s).

[3] UTI data is synchronised with an acoustic channel, and was collected for one participant in the spontaneous speech (SS) task, and all participants in the wordlist task. In addition there are acoustic-only recordings of both participants in the SS task.

Table 2 Overview showing the amounts of data recorded in each phase in minutes of speech, of Audio-only and UTI (including audio) data.

Type	**Corpus**	**Participants**	**Location**	**Style**	**No. of minutes (Audio)**	**No. of minutes (UTI)**
Scoping	n.a.	n=50-100 (age 5-50)	Various	live	n.a.	n.a.[4]
Pilot	n.a.	n=9 (6 males, 3 females, age 20-40)	Studio	mix	n.a.	37
Phase 1	WL07	n=14 (LM1-LM14, WC, age 12-13)	School	SS	322	n.a.
				WL	42	n.a.
Phase 2a		n=4 (LM: 1, 2, 7, 8)	School	SS	92	n.a.
				WL	64	n.a.
Phase 2b		n=10 (LM: 3, 4, 5, 6, 9, 10, 11, 12, 13, 14)	School	SS	230	33
				WL	15	9.5
Phase 2c		n=4 (LM: 1, 2, 3, 4)	Studio	SS	80	17
				WL	6	10
		Total		SS	12h 4mins	50
				WL	2h 7 mins	19.5
Phase 3	ECB08	n= 15 (7 males, 8 females, 8 WC, 7 MC, age 12-13)	Studio	SS	140	70.75
				WL	10	52.5
Post-hoc		4 (2 males, 2 females, MC, age 10-11)	Studio	SS	80	20.5
				WL	6	13
		Total		SS	3h 40 mins	91.25
				WL	16	65.5
The figures show minutes of recording per group in the audio-only and UTI conditions, for SS (spontaneous speech) and WL (wordlist) tasks. There are two different wordlists: Audio WL = a sociolinguistic wordlist, UTI WL = phonetic materials (see Annex 4).						

[4] No recordings were made. Live interaction was with hand-held probe.

Table 3 Speaker by speaker details of UTI data quantity showing subject identifiers for WL07, so that it is clear which subjects participated multiple times in which conditions. Only one speaker per pair was recorded in the SS condition

		Individual US minutes	
Type	Speaker	WL	SS[5]
Phase 2b	LM3	1	--[6]
	LM4	1	--
	LM5	--	--
	LM6	--	12.5
	LM9	--[7]	--
	LM10	4	4.5[8]
	LM11	1.5	--
	LM12	1	12
	LM13	0.5	--
	LM14	0.5	4
	Total	9.5	
Phase 2c	LM1	3.5	8.5
	LM2	2.5	--
	LM3	1.5	--
	LM4	2.5	8.5
	Total	10	17
	Total	19.5	31

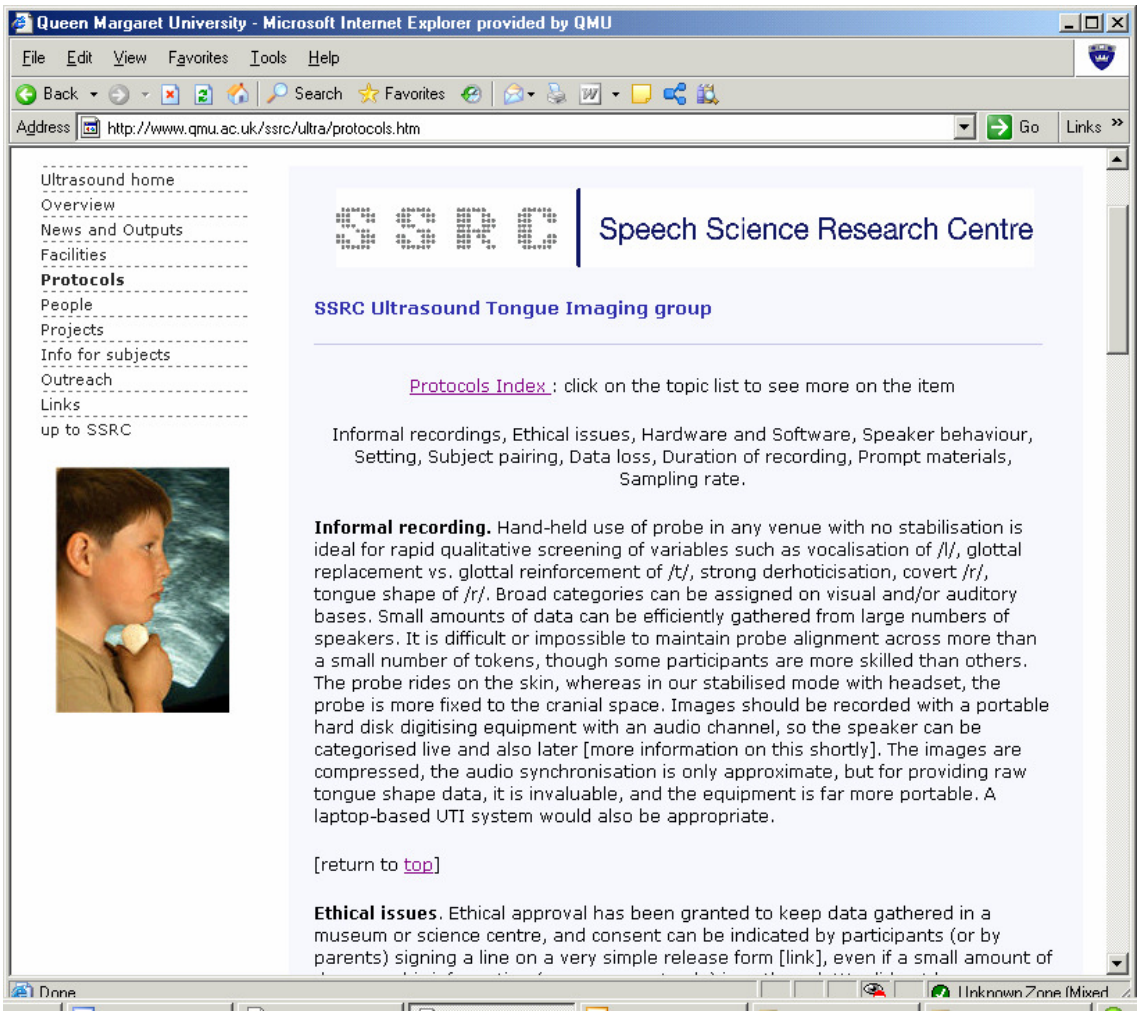
[5] Excluding samples recorded where researchers are talking for informants or fitting headsets.
[6] Spontaneous speech data not captured by software due to researcher error.
[7] Participant declined to be recorded for WL (wanted to get to the lunch queue early) after participating in SS condition as (unknowingly unrecorded) interlocutor.
[8] More recording time, but probe slips and data is unusable after recording 38.

Table 4 Speaker by speaker details for UTI data showing subject identifiers for ECB08 (including the extra data collected post-hoc from slightly younger MC children using the same methodology). Only one speaker per pair was recorded in the SS condition.

		Individual US minutes	
Type	Speaker	WL	SS
Phase 3 (WC)	LM15	3.5	8.5
	LM16	3.5	--
	LM17	3.5	10.5[9]
	LM18	3.5	--
	LF1	3.5	10
	LF2	3.5	--
	LF3	3.5	9.5
	LF4	3.5	--
	Total	28	38.5
Phase 3 (MC)	EM3	3.5	7
	EM4	3.5	--
	EM5	3.5	8.5
	EF3	3.5	7.5
	EF4	3.5	--
	EF5[10]	3.5	9.25
	EF6	3.5	--
	Total	24.5	32.25
	Total	52.5	70.75
Post hoc (MC)	EM1	3.5	7.5
	EM2	3.5	--
	EF1[11]	3.5	7+6
	EF2	3.5	--
	Total	14	20.5
	Total	66.5	91.25

[9] Informants knocked over amp, so there is no sound between 18-42. Extra recording time made up for this loss.
[10] Tongue image of EF6 (not included) until recording sample no. 7.
[11] Two separate recordings from 1-30 and from 31-55 – second recording contains little speech.

Annex 2
Contents page of SSRC/Ultra website page containing UTI research protocols



http://www.qmu.ac.uk/ultra/protocols.htm

Annex 3
Sociolinguistic word list

three
caught
think
weather
apple
hill
milk
butter
thought
people
feel
bottle
water
wine
bath
later
anything
batter
fatal
splinter
little
tooth
mouth
whine
why
bother
computer
Saturday
thanks
mother
watt
nothing
what
birthday
lock
thumb
teeth
seven
whether
other
with
either
apple
tree
path
hills
blether
something
smooth
breathe
brother
when
shelf
mat
sorry
white
pompom

Annex 4
Experimental Materials in ECB08

For the ECB08 Corpus, we focussed on materials which would be useful for the study of rhotic and derhoticised speech, because we expected only a small amount of derhoticisation. Since ultrasound is most effective when looking at slow moving articulations, we chose to focus on low back vowels. Mid and high vowels + /r/ result involve a fair amount of movement, and weak syllables tend to be short. The core of the design was to examine VL (a vowel plus a labial stop), VA (a vowel plus an alveolar stop), VrL and VrA. In addition, V and Vr were examined, even though the /r/-less cells are problematic: /□/ does not occur in open syllables at all, and /a/ is rare and has a front vowel quality, while all /ar/ are said to be low, back and long.

We did not have time to hide the materials among distractors. However, the participants reported nothing that let us think they understood that /r/ was the focus of the experiment. One participant suggested the words were being collected to make a story.

To allow for a warm-up, the first 10 words were not part of the design, but they did sample initial /r/, and /l/ in final position. This was followed by the 41 item word list (randomised), followed by a second repetition of the wordlist (in a different randomisation) and then 13 cool-down words (extended to 16 for the MC Edinburgh informants) to sample the entire vowel space.

The main materials sampled two vowels. As far as possible, they were placed in an open syllable, before a labial stop, before a coronal stop, and before the labial nasal. Real words were used with an intention that, with two repetitions of each, there would be 6 tokens per cell. Both voiced and voiceless stops were used to populate the design. With hindsight, the role of voicing in the stop following the /r/ is one of the more interesting factors. With more time it would have good to have /r/ before fricatives and data on /□/.

n=41	open	p	b	t	d	m
□	n.a.	pup	pub hub	hut butt	bud	bum mum hum
	purr fir fur	burp	(verb herb)[12]	hurt Burt	bird	(perm) firm
a/□	ma pa baa	map	-	hat	-	pam palm
	par bar far	harp parp	barb	heart part	hard	harm farm arm
weak			suburb		hammered	

Warm-up: ram, rum, lumber, lamb, cull, Mull, hulk, pill, cult, film, bulb
Cool-down: hem, beer [13], bear, beam, boom, hope, hip, for, awe, poor, fame, bore, hubbub, with extra cool-down materials for MC participants: sure, pure, bare.

[12] These bracketed items were pronounced with a different vowel from *burp* or *bird*, i.e. /□/, by some speakers but were used due to lack of suitable lexemes and because some speakers do not in fact use /□/.
[13] This item was mispronounced as /bir/ by a number of speakers so *bare* was added at a later stage to the end of the cool-down words.

Bibliography

Aitken, A.J. (1979). Scottish Speech: a historical view with special reference to the Standard English of Scotland. In *Languages of Scotland*. (Eds.) Aitken, A.J. & McArthur, T. London: Chambers. 85-118.

Byrd, Dani and Saltzman, Elliot (2003). The elastic phrase: modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics* (31) 149–180.

Delattre, Pierre and Freeman, Donald. C. (1968). A Dialect Study of American R's by x-ray motion picture. *Linguistics* (44) 29-68.

Dobson, Eric. J. (1957). *English Pronunciation*, *1500-1700*. Oxford: Clarendon Press.

Foulkes, Paul and Docherty, Gerard (2006). The Social Life of Phonetics and Phonology. *Journal of Phonetics* (43/4) 409–438.

Gick, Bryan (2002). The use of ultrasound for linguistic phonetic fieldwork. *Journal of the International Phonetic Association* (32/2) 3-17.

Gick, Bryan and Wilson, I. (2001). Pre-liquid excrescent schwa: what happens when vocalic targets conflict. *Proceedings of Eurospeech* 273-276.

Gick, Bryan, Kang, A.M., Whalen, D.H. (2002). MRI evidence for commonality in the post-oral articulations of English vowels and liquids. *Journal of Phonetics* (30) 357-371.

Gick, Bryan and Campbell, Fiona (2003). Intergestural Timing in English /r/. *Proceedings of the 15th International Congress of Phonetic Sciences* 1911-1914.

Gick, Bryan, Campbell, F., Oh, S., Tamburri-Watt, L. (2006). Toward universals in the gestural organization of syllables: A cross-linguistic study of liquids. *Journal of Phonetics* (34/1) 49-72.

Gordeeva, Olga B. and Scobbie, James M. (2007). Non-normative preaspirated voiceless fricatives in Scottish English: Phonetic and phonological characteristics. *QMU Speech Science Research Centre Working Paper WP12.*

Guenther, F. H., Espy-Wilson, C. Y., Boyce, S. E. Matthies, M. L. Zandipour, M., Perkell, J. S. (1999). Articulatory tradeoffs reduce acoustic variability during American English /r/ production. *Journal of the Acoustical Society of America* (105) 2854-2865.

Kerswill, Paul E. (1985). A sociophonetic study of connected speech processes in Cambridge English: An outline and some results. *Cambridge Papers in Phonetics & Experimental Linguistics* (4) 1–39.

Kerswill, Paul & Wright, Susan. (1990). On the limits of auditory transcription: a sociophonetic perspective. *Language Variation and Change* (2) 255-75.

Lawson, Eleanor (in preparation) Loquacity of participants during spontaneous discourse in an Ultrasound Tongue Imaging study. *QMU Speech Science Research Centre Working Paper WP14.*

Lawson, Eleanor, Stuart-Smith, J., and Scobbie, J.M. (in press) Articulatory insights into language variation and change: preliminary findings from an ultrasound study of derhoticisation. *Selected papers from NWAV 36, Pennsylvania Working Papers in Linguistics 14 (part 2).*

Lindau, Mona (1978). Vowel Features. *Language* (54) 541-563.

Lindau, Mona (1985). The story of /r/. In *Phonetic linguistics*. (Ed.) Fromkin, Victoria A. London: Academic Press. 157–168.

Local, John (2007). "Sound to sense: Introduction to the special session." *Proceedings of the 16th International Congress of Phonetic Sciences*. 182-183.

Mielke, Jeff, Baker, Adam and Archageli Diana (in press) Variability and homogeneity in American English /r/ allophony and /r/ retraction. In *Papers in Laboratory Phonology 10: Variation, Detail and Representation*. (Eds.) Fougeron, Cecile. & D'Imperio, Maria Paolo. Berlin: Mouton de Gruyter.

Oller, D. Kimbrough (1973). The effect of position in utterance on speech segment duration in English. *Journal of the Acoustical Society of America* (54/5) 1235 – 1247.

Romaine, Suzanne (1979). Postvocalic /r/ in Scottish English: Sound change in progress? In Sociolinguistic Patterns in British English. (Ed.) Trudgill, P. Edward Arnold: London. 145 – 157.

Scobbie, James (2007). "Biological and social grounding of phonology." *Proceedings of the 16th International Congress of Phonetic Sciences.* 225-228.

Scobbie, James (in preparation) invited newcastle

Scobbie, James, Stuart-Smith, J. & Sebregts, K. (2006). From subtle to gross variation : an ultrasound tongue imaging study of Dutch and Scottish /r/ (Poster presentation at Laboratory Phonology 10, Paris).

Scobbie, James M., Pouplier and Wrench,(2007). "Conditioning factors in external sandhi: An EPG study of English /l/ vocalisation. In *Proceedings of the 16th International Congress of Phonetic Sciences*. 441-444.

Scobbie, James Wrench and van der Linden (in preparation), (Proceedings of ISSP)

Scottish Executive (2006) *Scottish Index of Multiple Deprivation*. Edinburgh: Scottish Executive. Accessed April 2007 at: http://www.scotland.gov.uk/Resource/Doc/151578/0040731.pdf

Silverman, Daniel (1995). Phasing and Recoverability. New York: Garland.

Speitel, Hans-Henning & Johnston, Paul. (1983). ESRC End of Grant Report "A Sociolinguistic Investigation of Edinburgh Speech."

Sproat, Richard and Fujimura, Osamu (1993) Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics* (21) 291-311.

Stone, Maureen (1997). "Laboratory techniques for investigating speech articulation." In W. J.Hardcastle & J. Laver (Eds.) Handbook of Phonetic Sciences. Oxford: Blackwell. 11-32.

Stuart-Smith, Jane (2003) The phonology of Modern Urban Scots. In *The Edinburgh Companion to Scots*. (Eds. Corbett, J. McClure, D and Stuart-Smith, J. Edinburgh: Edinburgh University Press. 110-37.

Stuart-Smith, Jane (2005) *Is TV a contributory factor in accent change in adolescents?* Final Report to the ESRC. Grant No. R000239757.

Stuart-Smith, Jane (2007). A sociophonetic investigation of postvocalic /r/ in Glaswegian adolescents. *Proceedings of the 16th International Congress of Phonetic Sciences*. 1449-1452.

Stuart-Smith, Jane, Timmins, Claire and Tweedie, Fiona (2007). Talkin' Jockney: Accent change in Glaswegian. *Journal of Sociolinguistics* (11) 221-61.

Thomas, Erik R. (2002). Instrumental phonetics. In J.K. Chambers, Peter Trudgill and Natalie Schilling-Estes (Eds.) The Handbook of Language Variation and Change. Oxford; Blackwell Publishers Ltd. 168-200.

Wrench, A.A. and Scobbie, J.M. (2003). Categorising vocalisation of English /l/ using EPG, EMA and Ultrasound. In Sallyanne Palethorpe and Marija Tabain (Eds.) *Proceedings of the 6th International Seminar on Speech Production* (ISSP Sydney), 314-319.

Wright, Susan & Kerswill, P. (1989). Electropalatography in the analysis of connected speech processes. *Clinical Linguistics & Phonetics* (3) 49–57.