



10-1-2014

A mimicry study of adaptation towards socially-salient tongue shape variants

Eleanor Lawson
Eleanor.Lawson@glasgow.ac.uk

Jane Stuart-Smith
Jane.Stuart-Smith@glasgow.ac.uk

James M. Scobbie
jscobbie@qmu.ac.uk

A mimicry study of adaptation towards socially-salient tongue shape variants

Abstract

We know that fine phonetic variation is exploited by speakers to construct and index social identity (Hay and Drager 2007). Sociophonetic work to date has tended to focus on acoustic analysis, e.g. Docherty and Foulkes (1999); however, some aspects of speech production are not readily recoverable from an acoustic analysis. New articulatory analysis techniques, such as ultrasound tongue imaging (UTI), have helped to identify seemingly covert aspects of speech articulation, which pattern consistently with indexical factors, e.g. underlyingly, Scottish English middle-class and working-class coda /r/, have radically different tongue shapes and tongue gesture timings (Lawson, Scobbie and Stuart-Smith, 2011). This articulatory variation has gone unidentified, despite decades of auditory and acoustic analysis (Romaine, 1979; Speitel and Johnston, 1983; Stuart-Smith, 2007). UTI revealed that middle-class speakers tend to produce bunched variants of postvocalic /r/, while working-class speakers tend to produce tongue-tip raised variants (Lawson, Stuart-Smith and Scobbie 2011). We present the results of an ultrasound-based, speech-mimicry pilot study, which investigates how subtle articulatory information might be passed from speaker to speaker. Baseline articulatory information on /r/ was gathered for three Central-Scottish female, middle-class, pilot participants, who all used bunched /r/ variants in baseline. Participants mimicked audio-only stimuli, extracted from a socially-stratified audio-ultrasound corpus collected in Glasgow, Scotland. Analysis showed a range of mimicking behaviours from pilot participants including no modification from baseline, accurate discrimination between middle-class and working-class stimuli and adaptation from baseline, but failure to discriminate between middle-class and working class stimuli. This small-scale study provides new insights into the transmission of phonetic variation.

A mimicry study of adaptation towards socially-salient tongue shape variants

Eleanor Lawson, Jane Stuart-Smith and James M. Scobbie

1 Introduction

We know that fine phonetic variation is exploited by speakers to construct and index social identity (Hay and Drager 2007). Sociophonetic work to date has tended to focus on acoustic analysis, e.g. (Docherty and Foulkes 1999); however, some aspects of speech production are not readily recoverable from an acoustic analysis. One clear example of this type of “covert” variation is found in analyses of approximant /r/. Articulatory variation in approximant /r/, observed by researchers studying different varieties of English, has resulted in conflicting conclusions about the perceptual recoverability of the underlying tongue gestures. Studies of American English such as (Delattre and Freeman 1968, Guenther et al. 1999 and Mielke et al. 2010) have shown apparently idiosyncratic variation in the tongue shapes underlying approximant /r/, i.e. some speakers prefer to produce /r/ with their tongue tip raised and some with their tongue dorsum raised, or they show contextually-predictable variation. These findings support the notion of articulatory tradeoffs, where articulatory variation results from adjustments made in order to maintain a relatively stable acoustic signal.

Analysis of Scottish ultrasound tongue imaging (UTI) corpora has revealed that the postvocalic /r/ tongue shape is socially stratified. Lawson, Stuart-Smith and Scobbie’s articulatory analysis of the post-vocalic /r/s of fourteen eastern Central Scottish adolescents showed that middle-class and working-class speakers used radically different tongue shapes in the production of postvocalic /r/, tending towards *bunched* and *tip up* configurations respectively (Lawson et al. 2011); see Figure 1 below.

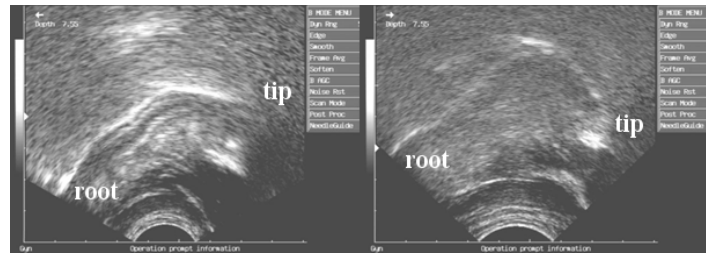


Figure 1: (Left) a midsagittal UTI still frame of a tongue-tip raised /r/ in *far*, produced by a working-class Scottish female. (Right) a midsagittal UTI still frame of a bunched /r/ in *far*, produced by a middle-class Scottish female. The tongue root is to the left side of the image and the tongue front is on the right side of the image.

Analysis of a second audio-ultrasound corpus of the speech of sixteen adolescents from western Central Scotland, collected in 2012, confirmed this pattern of stratification (Lawson et al. 2014). Although there was a continuum of variation between the “bookends” of dorsal bunched and retroflex /r/, a clear tendency was observed for working-class speakers in the eastern and western Central Belt to produce /r/ that involved tongue-tip raising to the back of the alveolar ridge, while the middle-class speakers from East and West tended to have a lowered tongue tip, retracted tongue root and the primary constriction formed at the palate rather than the alveolar ridge.

It was also found that tongue gesture timings for postvocalic /r/ were socially stratified, with middle-class speakers producing postvocalic /r/ with an early /r/ gesture, while working-class speakers often delayed the maximum of the /r/ gesture to a point after the offset of voicing meaning that /r/, although present, was often barely audible or inaudible (Lawson, Scobbie and Stuart-Smith, 2014); see also (Scobbie et al. 2009:15) for an instance of delayed, covert rhotic gestures in Dutch.

Social stratification of tongue shape in the production of /r/ implies that this aspect of articulation is perceptible or recoverable; in fact, the middle-class and working-class Scottish /r/ sound

different largely due to variation in the timing of the /r/ gesture, with the former having a strong rhotic quality, while the later is weakly rhotic (we use the descriptive term “derhoticised”) and often accompanied by pharyngealisation of the preceding vowel (Speitel and Johnston 1983:28). Indeed, it has been observed for many decades that Scottish postvocalic /r/ is socially stratified; see Stuart-Smith et al. (2014), or for over two centuries if we consider older discussions of stigmatised trilled/tapped versus approved and acceptable approximant /r/, e.g. Walker (1791).

Nevertheless, some of the articulatory variation revealed by recent UTI studies appears to have been happening under the radar as far as sociolinguists, dialectologists and phoneticians were concerned. For example, middle-class Scottish /r/ has traditionally been described as an *alveolar* or *retroflex* approximant, not a *bunched* approximant (Grant 1913, McAllister 1938, Romaine 1978, Stuart-Smith 2003). Grant (1913) and McAllister (1938) even provide midsagittal vocal-tract cutaways showing the alveolar and retroflex tongue shape involved in Scottish pronunciations of /r/. Likewise, although weak variants of /r/ had been observed in vernacular speech (Romaine 1978, Speitel and Johnston 1983 and Macafee 1983), it was speculated that the rhotic tongue gesture was underlyingly deleted, rather than delayed and inaudible. Johnston (1997), for example, speculated that postvocalic /r/ may be deleted at the articulatory level for these speakers, suggesting that the pharyngealisation of vowels before weak /r/ in working-class speech was a default voice-quality setting for the working-class Scottish social group and did not necessarily indicate the presence of an underlying tongue gesture related to /r/ (Johnston 1997:511).

Therefore we have a situation where speakers appear to be acquiring and using articulatory variants that sociolinguists, phoneticians and dialectologists did not know existed, even with detailed auditory and acoustic analysis. Several possibilities suggest themselves: (1) that speakers are picking up on visual cues, detecting tongue position and timing by studying one another’s mouths during conversation, (2) that speakers are reconstructing tongue configurations based on the coarticulatory effects of /r/ on surrounding sounds (for example, Lawson et al. (2013) have shown that postvocalic /r/ bunching is associated with a checked-vowel merger in middle-class Scottish English), and finally, (3) perhaps the most simple explanation is that there are acoustic cues for these articulatory variants that are very subtle (Jou and Espy-Wilson 2008), but nonetheless perceptible to speakers, all the more so when the variant in question is socially salient. See e.g. Lennon (2014), and Docherty and Foulkes (1999) for similar scenarios in Glaswegian and Newcastle and Derby speech respectively.

We aim to determine whether the articulatory variants of /r/ described above are perceptible from an audio-only stimulus by selecting tokens of words containing postvocalic /r/ with varying tongue configurations, presenting them aurally, randomized among distractors, to speakers who have been asked to mimic them, and then comparing the mean tongue shape of the mimicker’s habitual /r/ articulation to the mean tongue shape of /r/ produced in the two mimicry conditions. The results of this study will shed light on how speaker-listeners respond to auditory stimuli and provide more information on how adaptation towards another variety might take place.

2 Method

2.1 Stimuli collection

We recorded 16 participants aged 12–13 from Glasgow, forming a corpus we call WCB12. Half of those we recorded were male and half were female. Half were from a school in a middle-class area of the city and half were from a school in a working-class area of the city¹. Informants were recorded (audio and UTI) in an IAC sound-attenuated recording booth at the University of Glasgow in 2012. Participants were fitted with a stabilising headset (Scobbie et al. 2008) to hold the ultrasound probe and minimise its movement during the experiment, while also allowing the participant to freely move their head and upper body.

Traces of participants’ occlusal (bite) planes were obtained by asking them briefly to bite on a section of flat plastic and press their tongue against it. The probe angle was adjusted so that the occlusal plane appeared horizontal in the ultrasound image. Participants were recorded sipping

¹School statistics regarding percentages of top grades obtained, school leaver destinations and free school meals were available and confirmed socioeconomic trends in the two school catchments.

water at the beginning of the recording session, as filling the oral cavity with water obtains a trace of the participant's hard palate, which reveals the position of the tongue in relation to a stationary section of the vocal tract.

Participants were asked to read orthographic prompts presented to them on a monitor placed in front of them at eye level. Audio recordings were made using Beyer-Dynamic Opus 55 head-worn microphone. The recordings were sampled at 22kHz and a Mindray DP2200 ultrasound machine recorded ultrasound video at a target rate of 30 frames per second.

Although this method may seem invasive compared with most sociolinguistic data collection techniques, we were still able to gather examples of both strong and weak /r/ from participants, largely due to the fact that variation in /r/ is still below the level of conscious awareness and, unlike for example /t/-glottalling, is not generally remarked upon by Scottish English speakers. In fact, Scottish speakers when questioned show a surprising lack of awareness of the kind of /r/ they use and will for example state that Scottish people trill their /r/s, whereas trilled /r/ is now very rare in Scottish speech. The impact of the formal recording situation on /r/ appears to be minimal, see Lawson et al. (2008). So far we have encountered one case where an adolescent male participant produced trilled postvocalic /r/ while reading a word list, but not in his normal conversational speech; this speaker's data was not included in the study.

The audio recordings consisted entirely of CVrC nonwords such as *veerp* and *furf*², as an earlier pilot study had shown that the closest mimicry of the bunched tongue configuration occurred when the mimicker had not achieved lexical access (Lawson et al. 2011). Likewise, the only stimulus containing a derhoticised /r/ that was mimicked as /r/-less by the pilot mimicker was the sole token in the experiment that formed part of an /r/-ful, /r/-less minimal pair, namely 'hurt' and 'hut'. The pilot mimicker was presented with an audio recording of 'hurt', but produced 'hut'. Therefore it would seem that lexical access complicates interpretation of the phenomena that we are interested in. Episodic models of speech perception also suggest that low frequency words and nonsense words will be more closely mimicked (Goldinger 1998). Nonwords were therefore used as stimuli so that mimickers would not have lexical access, would not be able to guess whether the stimulus was intended to sound rhotic and would not be able to draw on a set of previously heard exemplars during production (except perhaps at the sublexical level, which is difficult to avoid).

It was also decided to focus on female speech due to our observation of greater social category differentiation in the WCB12 corpus between girls than between boys. We found that working-class girls were the strongest derhoticisers in our corpus, while middle-class girls were consistent bunchers; see Eckert (1989) for an early example of this gender-based pattern, and see also Stuart-Smith (2007).

Two main sets of rhotic stimuli were extracted from the WCB12 corpus; a set of 12 weakly /r/-ful, working-class stimuli and a set of 12 strongly /r/-ful, middle-class stimuli. The researchers aimed to select stimuli for each of these groups that were maximally different along three parameters: (1) tongue shape, (2) tongue gesture timing, and (3) audible rhoticity. However, we will only consider tongue shape in our analysis at this time.

For parameter 1, video files from the stimulus corpus were inspected in order to identify examples of unambiguous bunched or tip-up /r/ articulations, only bunched /r/ articulations were selected for the middle-class section of the stimuli and only tip-up or front-up articulations were selected for the working-class section of the stimuli.

For parameter 2, the normalised duration between the offset of voicing and the maximum of the /r/ gesture was determined (called "normalised lag" in Table 1 below). Potential stimulus examples were identified from the most extreme positive or negative lag values. All working-class stimuli selected had an /r/ gesture with a positive lag (mean 0.5), meaning that the /r/ gesture maximum occurred during the production of the following labial consonant. All middle-class stimuli selected had an /r/ gesture with a negative lag (mean -0.34), meaning that the /r/ gesture maximum occurred before the onset of the closing labial consonant.

For parameter 3, the three authors carried out an auditory rating task using a Praat multiple

²The vowel element in the nonword stimuli was limited to four vowels dispersed around the vowel space /i/, /o/, /a/ and /ʌ/ in order to limit the number of factors we needed to take into account during analysis. Flanking consonants were always labials or /h/ in order to avoid potential coarticulatory effects of consonants on /r/.

forced choice interface that randomised the presentation of stimuli and allowed them to be rated on a rhoticity index of 1 – 5, (1= no audible /r/ and 5=strong rhotic quality); see Lawson et al. (2011) for more information on this rhoticity index. A mean /r/ index score was then assigned to each /r/-ful item in the corpus to aid stimulus selection; see the second and sixth columns of Table 1. The selected working-class /r/ had a mean rhoticity index score of 2.0 and the selected middle-class /r/ stimuli had a mean index score of 4.2.

A final consideration was to try to avoid always using the same speaker for weak and strong rhotic variants in order to avoid participants identifying particular speakers during the task and associating them with particular pronunciations of /r/.

By taking into account three key parameters of tongue shape, tongue gesture timing and audible rhoticity, a set of 24 audio recordings (12 with working-class /r/, 12 with middle-class /r/) were exported from the WCB12 corpus to be played to informants who were asked to “mimic them as closely as possible”. The stimulus words that were selected, along with their speaker code, can be seen in the first and fourth columns of Table 1. In addition to CVrC, words CVC equivalents were added (as long as they were nonwords) in order to prevent participants from always expecting r+C coda consonant clusters. Other items from the word list were included as distractors (see Table 2); some of these were nonsense words deliberately included as orthographic prompts in the stimulus word list and others were serendipitous mispronunciations where participants metathesized consonant and vowel sequences, deleted segments and occasionally produced odd, intermediate speech sounds containing features indicative of more than one consonant sound, or with a vowel quality intermediate between two canonical Scottish vowels. We made use of these to augment the distractor list, to distract from the focus of the study and to force participants to focus on mimicking phonetic quality, rather than attempting a phonological parse.

Finally, in order to test the theory that coarticulatory cues could be involved in the perception of tongue configuration, at the end of the mimicry task informants were instructed to listen to two sequences of four words *meerp*, *marp*, *moarp*, *murp* produced by a middle-class and a working-class speaker. The informants were instructed to mimic only the last word, containing the [ʌ] vowel which is particularly strongly modified by different types of following /r/ in Scottish English; see Lawson et al. (2013). We hypothesized, therefore, that hearing words where /r/ occurred in different vowel environments in close temporal proximity, as might happen during a conversation with another speaker, might improve the chances of the listener being able to reconstruct and reproduce the correct /r/ variant.

All in all, there were 24 target /r/-ful words and 58 distractors (see Table 1). The mean intensity of recordings used as stimuli were scaled in order to avoid variation in loudness from token to token. A different randomization of the word list was used for each member of each block of three mimicry participants.

Informants were asked at the end of the study whether they had noticed any particular focus to the stimuli, and some of them said that they had not been aware that the study was focussing on any particular sound. Others said they thought there was a focus on “vowels”, or the particular vowels /i/ or /ʌ/, or the consonants /p/ and /f/. However, one participant said “maybe r”, and another said that she thought there was a focus on “Scottish sounds like /x/ and... ear”³. The authors assume that in saying the word “ear” and identifying it as a “Scottish sound” the participant was identifying rhoticity a focus for the study.

³N.B. The participant was aware that we had asked schools for informants with two Scottish parents and that we were interested in Scottish speech. In fact, only one of the stimulus words had /x/ in it.

Weak postvocalic /r/ N=12	Mean /r/ index score	Normalised gesture lag	Strong postvocalic /r/ N=12	Mean /r/ index score	Normalised gesture lag	Non /r/ minimal pairs N=16	
GWF1_hurp	1.7	0.89	GMF4_hurp	4.7	-0.89	GMF1_hup	GWF4_hup
GWF1_mar_p	1.7	0.69	GMF2_mar_p	3.3	-0.25	---	
GWF1_vurf	1.7	1.15	GMF4_vurf	5.0	-0.63	GMF3_fuff	GWF2_fuff
GWF3_murp	1.7	0.50	GMF1_murp	5.0	-0.38	GMF4_mup	GWF1_mup
GWF3_veerp	1.7	0.52	GMF1_veerp	4.3	-0.33	GMF4_veep	GWF3_veep
GWF1_hoarf	3.0	0.36	GMF4_hoarf	4.3	-0.25	GMF3_hoaf	GWF3_hoaf
GWF4_arf	2.0	0.03	GMF1_arf	4.0	-0.26	---	
GWF4_varp	2.3	0.00	GMF2_varp	3.7	-0.22	GMF3_vap	GWF3_vap
GWF4_meerf	1.3	0.48	GMF1_meerf	3.7	-0.26	GMF3_meef	GWF4_meef
GWF1_boarp	2.0	0.48	GMF1_boarp	4.7	-0.21	GMF1_bope	GWF1_bope
GWF1_peerf	2.3	0.46	GMF2_peerf	4.3	-0.21	---	
GWF1_moarp	2.7	0.47	GMF2_moarp	3.7	-0.14	---	
Mean	2.0	0.50	Mean	4.2	-0.34		

Table 1: Stimulus items with /r/ presented to participants for mimicry, total N=40, along with mean index score and temporal lag for /r/, and corresponding minimal pairs without /r/.

Distractors N=42					
GMF1_glope	GMF3_splulf	GWF2_smap	GWF4_kulk	GWF4_fwaf	GWF1_splump
GMF1_hap	GWF1_meef	GWF2_sproke	GWF4_spluff	GWF4_gloop	GMF1_slup
GMF1_tweemp	GWF1_smeelf	GWF3_barsp	GMF3_mump	GWF3_uck	GMF2_eeb
GMF2_ooch	GWF2_brasp	GWF3_meefl	GWF2_xorf	GMF2_fap	GMF3_smelf
GMF2_smeel	GWF2_glope	GWF4_fuff	GMF1_bolp	GMF3_paa	GMF1_rutchill
GMF2_tweemp	GWF2_kluff	GMF3_borap	GWF1_oach	GWF4_hof	GWF2_fuff
GMF3_kulff	GWF2_slup	GWF4_glip	GMF2_smu	GWF2_tweeb	GWF4_spleemf

Table 2: Non-/r/ stimulus items presented to participants for mimicry, total N=42

2.2 Mimicry study participants

Nine mimicry participants were recruited; all female and aged between 13 and 21, all were from the Scottish Central Belt, and all but one were from Glasgow (see Table 3).

There were two recording phases, an initial pilot phase using older female participants in their late teens and early twenties, who were recorded at Queen Margaret University, Edinburgh, (QMU), and then the main recording phase where six participants in their early teens were recorded at the University of Glasgow (GU).

Recording location	Pilot group ⁴	Working-class group	Middle-class group
QMU	Pilot 3 West Lothian, age 18	GWF5, Glasgow, age 13	GMF5, Glasgow, age 13
GU	Pilot 5 Glasgow, age 21	GWF6, Glasgow, age 13	GMF6, Glasgow, age 13
GU	Pilot 6 Glasgow, age 19	GWF7, Glasgow, age 14	GMF7, Glasgow, age 13

Table 3: Mimicry participants’ demographic details. QMU = Queen Margaret University, Edinburgh. GU = University of Glasgow.

⁴Pilot participant numbering begins at 3 due to an earlier pilot study based on real-word mimicry that involved two male speakers. Pilot 4 in the present pilot study was found to have a hearing impairment which affected her ability to accurately identify sounds in the audio stimuli, and she was therefore excluded from the study.

2.2.1 Baseline recordings

Participants were first recorded (with audio and ultrasound) reading a randomized word list aloud that contained real words; 23 monosyllabic words containing stressed postvocalic /r/ e.g. *beer*, *arm*, *burp*, *oar*, and 55 distractors. These postvocalic /r/ words contained one of the four prerhotic vowels found in the mimicry stimuli /i/, /a/, /ʌ/ and /o/ to match the vowel conditions in the mimicry section of the study. Orthographic stimuli were presented one at a time on a screen in front of the participant, who was again fitted with the ultrasound headset and probe. This word list provided information on the speakers' *baseline* tongue configuration during the production of postvocalic /r/. During the mimicry phase of the study, participants were instructed to listen to the audio stimuli presented to them one at a time through headphones and then to "copy what they heard and how it was said as closely as possible, as if they were an echo." No orthographic prompts were shown during the mimicry phase of the recording session. The ultrasound recording began as the prompt started playing, so that participants could begin mimicking immediately, in order to avoid participants having to hold audio traces in their working memory for any length of time – which we hypothesized could lead to normalization towards habitual articulatory strategies. However, we did not instruct participants to mimic the stimulus as fast as possible.

3 Analysis

3.1 Tongue-shape analysis

Analysis of tongue shape was carried out semi-automatically using Articulate Assistant Advanced (AAA) software (Wrench 2012). A contour line, henceforth a *spline*, was fitted to the midsagittal tongue-surface (see Figure 3) at the temporal point of maximum constriction for /r/. This process involves selecting the video frame that corresponds to the point of maximum constriction. Boundaries are manually set (Figure 3:2) between which an algorithm searches for a bright surface (Figure 3:3). Thereafter, any part of the spline that does not follow the visible tongue surface is manually corrected and sections of spline that do not correspond to a tongue surface, e.g. beyond the tip of the tongue, are made invisible by lowering the confidence value of the spline (Figure 3:4). A spline was also manually fitted to the hard palate and alveolar ridge and the occlusal plane (Figure 3:4), which function as reference points within the vocal tract.

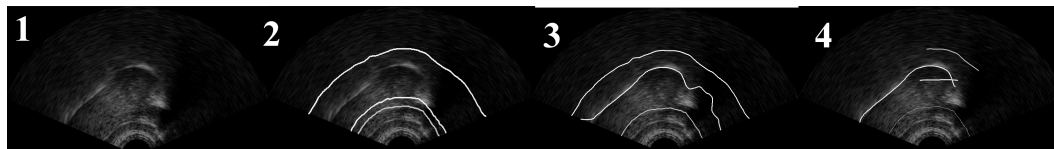


Figure 2: Semi-automatic spline fitting process from left to right: (1) before spline fitting, (2) parameters set (upper and lower white lines) for edge-detection algorithm, (3) after automatic edge detection, (4) after manual adjustment of the spline and with palate trace and bite plane added.

The tongue splines, bite plane and palate splines were then extracted to a work space for comparison. Around 47 splines were obtained per participant: 23 /r/ splines in the baseline condition and 12 /r/ splines in each of the mimicry conditions. Three mean tongue-surface splines were created for each of the three experimental conditions: *baseline*, *working-class mimic (WC mimic)* and *middle-class mimic (MC mimic)*.

4 Results

4.1 Tongue shape modification

Table 4 below shows how each of the nine participants modified their tongue shape from baseline (black line) to when they were mimicking working-class stimuli (dark grey) and middle-class stimuli (light grey). The uppermost black line in each cell is the hard-palate trace. The quasi-

horizontal dotted line represents the participants' occlusal (bite) plane. The tip of the tongue is found to the right of each cell, and the root of the tongue is to the left of each cell.

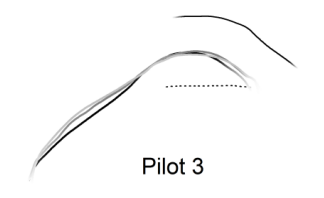
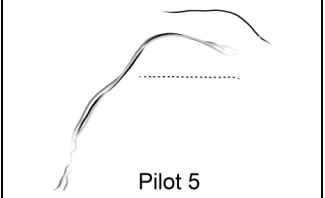
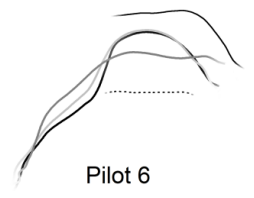
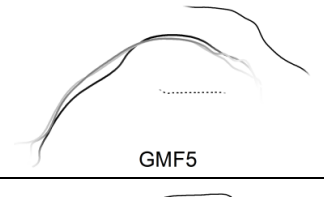
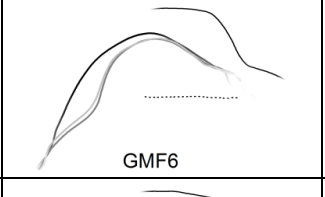
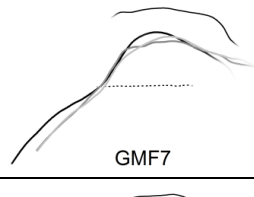
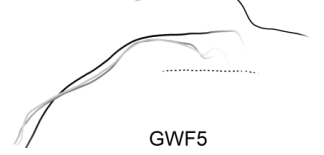
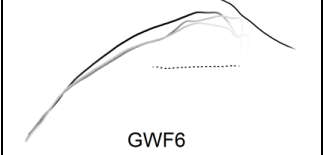

Pilot participants	 Pilot 3	 Pilot 5	 Pilot 6
Middle-class participants	 GMF5	 GMF6	 GMF7
Working-class participants	 GWF5	 GWF6	 GWF7

Table 4: Midsagittal tongue configurations in *baseline* (black); *working class mimic* (dark grey) and *middle class mimic* (light grey) conditions. The uppermost black line represents the hard palate. The dotted line marks a trace of the occlusal (bite) plane of the participant.

For the middle-class participants that comprise the top two rows of the table (pilot and middle-class informants), we might have expected little variation between their *baseline* tongue shape and the tongue shape produced in the *MC mimic* condition. On the other hand, in the *WC mimic* condition, we might expect to see the middle-class participants using a less bunched, more tip-up tongue configuration. In fact, it appears that only one of the middle-class participants, Pilot 6, successfully modified her tongue configuration towards that of the working-class tip-raised /r/ stimuli. Although working-class and middle-class stimuli were randomized throughout the experiment, Pilot 6 always produced the appropriate bunched or tip-raised variant depending on whether the stimulus was produced by a working-class or a middle-class speaker. Reporting results for the other two participants in the Pilot group is equally uncomplicated, as these participants did not deviate from their baseline production of /r/ in either mimicry condition.

For the younger middle-class participants in row two of Table 4, the results are more complex. GMF5 produced a front-bunched /r/ variant in *baseline*. In the mimicry condition, she produced variants with less of a dip in the region behind the tongue dorsum, but maintained the lowered tongue tip in her production of /r/ and she did not differentiate between the *WC mimic* and *MC mimic* conditions.

GMF6 produced a bunched /r/ in *baseline* that had no dip in the region behind the tongue dorsum, and in the mimic conditions, produced a different kind of bunched /r/, namely one with a post-dorsal dip, but again she did not differentiate between the *WC mimic* and *MC mimic* conditions, producing the same kind of modifications when mimicking both types of stimuli.

Interpreting the results for GMF7 are difficult as the headset was removed between *baseline* and *WC/MC mimic* recording sessions. The two sets of data were re-orientated according to the palate trace and occlusal-plane trace recorded at the beginning of each session; however the front- and higher position of the tongue root in *WC/MC mimic* condition might suggest that reorientation was not completely successful. If we take this possible issue of unsuccessful reorientation into account, it does not appear that this participant deviated much from her baseline tongue configuration in either the *WC mimic* or *MC mimic* conditions.

For the working-class participants found in the bottom row of Table 4, we might have expected little variation between their *baseline* tongue shape and the tongue shape produced in the

WC mimic condition. When mimicking middle-class stimuli, we might have expected to see the working-class participants producing bunched tongue configurations with a more lowered tongue tip, perhaps a dip in the region behind the tongue dorsum and a more retracted tongue root, or we might have expected to find some other strategy for producing a strongly rhotic approximant /r/ variant, e.g. strong retroflexion.

We can see that subtle modification towards a bunched /r/ occurred for two out of the three informants. GWF5 modified from her baseline tip-up /r/ towards a variant with a more retracted tongue root, lowered tongue tip and a slight post-dorsal dip; however, she did not differentiate between the *WC mimic* and *MC mimic* conditions, producing something more bunched on average when mimicking both kinds of stimuli.

GWF7, on the other hand, was the one other participant who appeared to have been able to successfully differentiate between working-class and middle-class stimuli; while her tongue shape in the *WC mimic* condition matches her own *baseline* tip-up tongue contour, the tongue contour in the *MC mimic* condition has much more of a bunched tongue configuration with a slightly lowered tip and a dip in the region behind the tongue dorsum.

GWF6 did not deviate from her *baseline* tip-up tongue configuration.

On five occasions, derhoticised working-class stimuli were mimicked as /r/-less, but by working-class participants GWF5 and GWF7. None of the middle-class participants mimicked /r/-ful stimuli as /r/-less. The following section discusses the implications of these results.

5 Discussion

The results of the tongue shape modification analysis provide some insights into how articulatory adaptation occurs. Not all speakers adapt towards the underlying tongue shape of the stimulus /r/. We also see that, in this study, adaptation is rarely discrete, but rather appears most often to involve blending and perseveration effects; that is to say, only two participants consistently differentiated between the tongue shapes of the working-class and middle-class stimuli. We also saw a few instances of misperception or misinterpretation of the stimulus where stimuli with late, masked /r/ gestures were mimicked as /r/-less.

To answer the question set out in the introduction: are the bunched and tip up variants of /r/ perceptible from an audio-only stimulus? It would appear that the bunched and tip-up tongue configurations can be perceived or recovered by some of the speakers in our study when presented with only an audio stimulus, as we see examples of some of the mimicry participants adapting to both of these tongue shapes while mimicking the audio stimuli. We found no instances of participants attempting to mimic a bunched /r/ variant by, for example, producing extreme retroflexion to match the strong rhotic quality of the bunched /r/. Therefore it would seem that the articulatory tradeoffs argument, where speakers use different articulatory strategies to maintain a stable acoustic signal, does not hold in this particular instance. However, the misperception of derhoticised stimuli in the copying of /r/-ful stimuli as /r/-less suggests that the acoustic cues in some of the derhoticised stimuli were subtle enough to be missed, rendering the /r/ articulations truly covert. This misperception offers a possible mechanism for postvocalic /r/-loss in Scottish English; however, it should also be noted that the majority of the derhoticised /r/ stimuli were mimicked as /r/-ful.

There was no observable ameliorative effect of providing a greater range of vowel environments in the audio stimulus to provide a greater range of coarticulatory cues. Those who did not copy the underlying tongue shape of the stimulus /r/ when they heard one nonsense word did not perform better when they heard four nonsense words, containing four different vowels, played in sequence. This result suggests that some of the mimickers in our study were able to pick up on subtle cues in the acoustic signal rather than reconstructing the tongue shape based on coarticulatory information.

One of the main findings of this study was that there was such a variety of individual mimicry strategies. While four participants did not deviate from their baseline tongue configuration, the remaining five participants adjusted their tongue shapes in various ways in the mimicry condition. We cannot assume that the four participants who did not alter their tongue shape from baseline did not perceive the tongue shape variation or were incapable of copying it. It might be that other aspects of the audio stimulus were more salient to them or they chose to copy other aspects of the

stimulus. All participants made some attempt to copy the stimulus, e.g. the breathy voice quality of a speaker, the pitch of the utterance, intonation pattern or the quality of the vowel.

Only two participants, Pilot 6 and GWF7, successfully discriminated between tongue shapes in each of the mimicry conditions enough to produce mean tongue shapes for the two mimicry conditions that looked distinct from one another. One of these two mimickers, GWF7, identified /r/ as the focus of the study. The most successful /r/ mimicker, Pilot 6, on the other hand, reported that she did not notice any particular focus to the study. One of the Pilot participants, Pilot 5, had reported “trying to speak like a ned⁵” as a young adolescent and mentioned that her high school had a mixed catchment area in terms of socioeconomic classes. It might have been expected that this speaker would be the one to adapt her tongue shape to that of the working-class stimuli; nevertheless, she did not adapt her tongue shape from baseline. Pilot 6, on the other hand, reported living in a very middle-class area, that her friends were middle-class and also reported that she did not have any aptitude for language learning, and yet she was the most successful participant at discriminating between tongue shapes in the mimicry condition, based on the audio stimuli.

For the remaining three participants who deviated from their baseline tongue shape, GMF5, GMF6 and GWF5, the mimicking of audio stimuli appeared to involve continued influence of previously-heard stimuli, rather than a discrete adaptation to each individual stimulus. For these three participants, the mean tongue shapes produced under the two mimicry conditions look different from the baseline tongue shapes, but identical to one another. For two out of these three participants (GMF6 and GWF5), the adaptation was in the direction of bunched /r/, perhaps because it is an auditorily stronger rhotic than that produced by working-class speakers; even though GMF6 produced a bunched /r/ in *baseline* condition, she produced a *different* bunched variant, with a more obvious dip in the region behind the tongue dorsum, in both mimicry conditions. GMF5 on the other hand continued to produce a bunched /r/ in the mimicry condition, but produced one with a less obvious postdorsal dip. The fact that there was deviation from *baseline*, but no clear differentiation between the mimicry conditions, could support an episodic model of perception-production (Goldinger 1998), in that recently-heard exemplars continue to have an effect on production.

This study is clearly small-scale; however, it adds to what we already know about the social stratification of bunched and tip-up /r/ (Lawson et al. 2011) and derhoticised /r/ in Scottish English, (Lawson et al. 2014), confirming that speakers from Central Belt communities can be sensitive to the subtle acoustic cues resulting from underlying articulatory variation and that they are able to use these cues to copy the tongue gestures produced by other speakers from their community.

References

- Delattre, Pierre and Freeman, Donald C. 1968. A dialect study of American r's by x-ray motion picture. *Linguistics*, 6:29–68.
- Docherty, Gerard and Foulkes, Paul. 1999. Newcastle upon Tyne and Derby: instrumental phonetics and variationist studies. In *Urban Voices: Accent Studies in the British Isles*, ed. P. Foulkes and Docherty, G.J., 47–71. London: Arnold.
- Eckert, Penelope. 1989. The Whole Woman: Sex and Gender Differences in Variation. *Language Variation and Change*, 1:245–269.
- Goldinger, Stephen. 1998. Echoes of Echoes? An Episodic Theory of Lexical Access. *Psychological Review*, 105:251–279.
- Grant, William. 1913. *The Pronunciation of English in Scotland*. Cambridge: Cambridge University Press.
- Guenther, Frank H., Espy-Wilson, Carol Y., Boyce, Suzanne E., Matthies, Melanie L., Zandipour, Majid and Perkell, Joseph.S. 1999. Articulatory tradeoffs reduce acoustic variability during American English /r/ production. *Journal of the Acoustical Society of America*, 105:2854–2865.
- Hay, Jennifer and Drager, Katie. 2007. Sociophonetics. *Annual Review of Anthropology*, 36:89–103.
- Johnston, Paul. 1997. Regional Variation. In *The Edinburgh History of the Scots Language*, ed. Jones, C., 443–513. Edinburgh: Edinburgh University Press.

⁵A derogatory Scottish slang term for a lower-class person, usually a young person, with connotations of lack of class and involvement in antisocial behaviour.

- Jou, Xinhui. and Espy-Wilson, Carol Y. 2008. A magnetic resonance imaging-based articulatory and acoustic study of “retroflex” and “bunched” American English /r/. *Journal of the Acoustical Society of America* 23:4466–4481.
- Lawson, Eleanor, Scobbie, James M. and Stuart-Smith, Jane. 2014. A socio-articulatory study of Scottish rhoticity. In *Sociolinguistics in Scotland*. ed., Lawson, R., 53–78. London: Palgrave Macmillan.
- Lawson, Eleanor, Scobbie, James M. and Stuart-Smith, Jane. 2011. A single-case study of articulatory adaptation during acoustic mimicry. In *Proceedings of the 17th international congress of phonetic sciences*.
- Lawson, Eleanor, Stuart-Smith, Jane, and Scobbie, James M. 2008. Articulatory Insights into Language Variation and Change: Preliminary Findings from an Ultrasound Study of Derhoticization in Scottish English. In *U. Penn Working Papers in Linguistics 8.3: Selected Papers from NWAV 30*, pp. 102–110.
- Lawson, Eleanor, Scobbie, James M. and Stuart-Smith, Jane. 2011. The social stratification of tongue shape for postvocalic /r/ in Scottish English1. *Journal of Sociolinguistics*, 15:256–268.
- Lawson, Eleanor, Scobbie, James M. and Stuart-Smith, Jane. 2013. Bunched /r/ promotes vowel merger to schwar: An ultrasound tongue imaging study of Scottish sociophonetic variation. *Journal of Phonetics*, 41:198–210.
- Lennon, Robert. 2014. Increased exposure can aid perception of ambiguous /r/ variants in Glasgow. Paper presented at the 2014 BAAP Colloquium, Oxford.
- Macafee, Caroline. 1983. *Glasgow. Varieties of English Around the World. Text series T3*. Amsterdam: Benjamins.
- McAllister, Anne. 1938. *A Year's Course in Speech Training*. London: University of London Press Ltd.
- Mielke, Jeff, Baker, Adam and Archangeli, Diana. 2010. Variability and homogeneity in American English /r/ allophony and /s/ retraction. Variation, Detail, and Representation. Paper presented Laboratory Phonology 10, Paris, France.
- Romaine, Suzanne. 1978. Postvocalic /r/ in Scottish English: Sound change in progress. In *Sociolinguistic Patterns in British English*, ed. Trudgill, P., 144–158.
- Scobbie, James M., Sebregts, Koen and Stuart-Smith, Jane. 2009. Dutch rhotic allophony, coda weakening, and the phonetics-phonology interface. In *Queen Margaret University Working Papers*, Edinburgh.
- Scobbie, James M., Wrench, Alan A. and van der Linden, Marietta. 2008. Head-probe stabilisation in ultrasound tongue imaging using a headset to permit natural head movement. In *Proceedings of the Eighth International Seminar on Speech Production*.
- Speitel, Hans H. and Johnston, Paul A., 1983. *A Sociolinguistic Investigation of Edinburgh Speech*. Economic and Social Research Council.
- Stuart-Smith, Jane. 2003. The phonology of modern urban Scots. In *The Edinburgh Companion to Scots*, eds. Corbett, J., McClure, J.D. and Stuart-Smith, J., 110-137. Edinburgh, U.K.: Edinburgh University Press.
- Stuart-Smith, Jane. 2007. Empirical evidence for gendered speech production: /s/ in Glaswegian. In *Change in Phonology (Laboratory Phonology 9)*, eds. Cole, J. and Hualde, J., 65-86. Berlin: Mouton de Gruyter.
- Stuart-Smith, Jane., Lawson, Eleanor. and Scobbie, James M. 2014. Derhoticisation in Scottish English: a sociophonetic journey. In *Advances in Sociophonetics*, eds. Celata, C. and Calamai, S. Amsterdam: John Benjamins.
- Walker, John. 1791. *A Critical Pronouncing Dictionary*. London: G. G. J. & J. Robinson.
- Wrench, Alan. 2012. *Articulate Assistant Advanced User Guide*. Edinburgh: Articulate Instruments Ltd.

Eleanor Lawson
 James M. Scobbie
 CASL research centre
 Queen Margaret University, Edinburgh, EH21 6UU
 elawson@gmu.ac.uk
 jscobbie@gmu.ac.uk

Jane Stuart-Smith
 English Language/Glasgow University Laboratory of Phonetics (GULP)
 University of Glasgow, Glasgow, G12 8QQ
 Jane.Stuart-Smith@glasgow.ac.uk