

MARCH 26 2018

The role of gesture delay in coda /r/ weakening: An articulatory, auditory and acoustic study

Eleanor Lawson; Jane Stuart-Smith; James M. Scobbie



J. Acoust. Soc. Am. 143, 1646–1657 (2018)

<https://doi.org/10.1121/1.5027833>



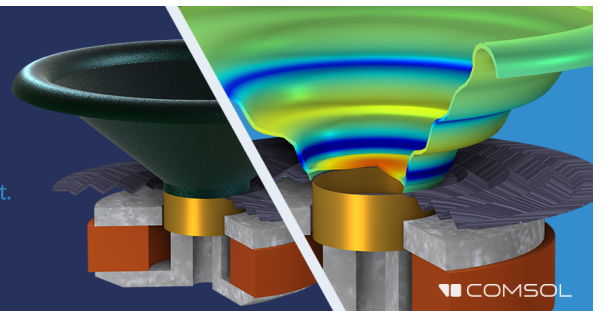
CrossMark



Take the Lead in Acoustics

The ability to account for coupled physics phenomena lets you predict, optimize, and virtually test a design under real-world conditions – even before a first prototype is built.

» Learn more about COMSOL Multiphysics®



COMSOL

The role of gesture delay in coda /r/ weakening: An articulatory, auditory and acoustic study

Eleanor Lawson,^{1,a)} Jane Stuart-Smith,² and James M. Scobbie¹

¹CASL Research Centre, Speech and Hearing Sciences, Queen Margaret University, Edinburgh, EH21 6UU, United Kingdom

²GULP Lab, English Language, 12, University Gardens, University of Glasgow, Glasgow, G12 8QQ, United Kingdom

(Received 5 June 2017; revised 23 February 2018; accepted 6 March 2018; published online 26 March 2018)

The cross-linguistic tendency of coda consonants to weaken, vocalize, or be deleted is shown to have a phonetic basis, resulting from gesture reduction, or variation in gesture timing. This study investigates the effects of the timing of the anterior tongue gesture for coda /r/ on acoustics and perceived strength of rhoticity, making use of two sociolects of Central Scotland (working- and middle-class) where coda /r/ is weakening and strengthening, respectively. Previous articulatory analysis revealed a strong tendency for these sociolects to use different coda /r/ tongue configurations—working- and middle-class speakers tend to use tip/front raised and bunched variants, respectively; however, this finding does not explain working-class /r/ weakening. A correlational analysis in the current study showed a robust relationship between anterior lingual gesture timing, $F3$, and percept of rhoticity. A linear mixed effects regression analysis showed that both speaker social class and linguistic factors (word structure and the checked/unchecked status of the prerhotic vowel) had significant effects on tongue gesture timing and formant values. This study provides further evidence that gesture delay can be a phonetic mechanism for coda rhotic weakening and apparent loss, but social class emerges as the dominant factor driving lingual gesture timing variation.

© 2018 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.1121/1.5027833>

[BVT]

Pages: 1646–1657

I. INTRODUCTION

A. Phonetic basis for coda consonant weakening

The notion that there is a phonetic basis for the cross-linguistic tendency of coda consonants to lenite, vocalize, or be deleted has, for a long time, been of interest to phoneticians and phonologists. Articulatory studies of speech have identified two key mechanisms that can underlie coda consonant weakening: articulatory reduction—a decrease in the magnitude of articulatory gestures—and variation in gesture timing—change in the timing of articulatory gestures relative to other speech events.

Browman and Goldstein's (1995) x ray microbeam study of American English consonants observed gestural reduction by tracking the position of pellets attached to the tongue blade and showed a reduction in height when /l/, /t/, and /n/ were in coda position, in comparison to onset position. They also observed a reduction in lip constriction for /p/ when it was in coda rather than onset position. They suggested that this phenomenon might be caused by a general reduction in speaking effort over the time course of the syllable (Browman and Goldstein, 1995).

A large body of articulatory research, much of it focused on /l/, uncovered another mechanism that potentially

contributes to coda consonant weakening, namely, variation in the synchronicity of the primary and secondary lingual gestures (Sproat and Fujimura, 1993; Recasens and Farnetani, 1994; Krakow, 1989). Sproat and Fujimura's study of American English /l/, using x ray microbeam, found that coda /l/ darkness correlated strongly with the acoustically measured duration of the rime containing the coda /l/. It was found that the stronger the following phonological boundary, the longer the preceding syllable rime and the darker the /l/, quantified in terms of $F2-F1$. Sproat and Fujimura initially hypothesised shorter rimes led to articulatory undershoot, whereby the tongue dorsum failed to retract as much as it would be able to in a longer rime; however, it was also found that longer rimes were associated with a greater durational difference between the primary apical gesture and the secondary dorsal gesture for /l/. In other words, both gestural undershoot and increased dissociation of gestures contributed to a darker, more vocalised, quality. The boundary that conditioned both the greatest degree of tongue dorsum retraction and the greatest degree of temporal separation between apical and dorsal gestures was the “major intonation boundary,” i.e., /l/ in pre-pausal position.

The possibility that variation in gesture synchrony could account for more than changes in phonetic quality, i.e., could be a mechanism for diachronic segment deletion, has been suggested by Recasens and Farnetani (1994), who carried out an electropalatography (EPG) study of contact patterns for

^{a)}Electronic mail: elawson@qmu.ac.uk

phrase-initial and phrase-final /l/, produced by single speakers of American English, Catalan, and Italian. They noted that the alveolar gesture of phrase-final dark /l/ in Catalan and American English not only occurred later than the dorsal gesture, but was found to occur partially or completely after the offset of voicing, leading to apparent consonant deletion at the auditory and acoustic levels, but not at the articulatory level, i. e., an apical articulatory gesture was present that was “devoid of acoustic consequences” (Recasens and Farnetani, 1994, p. 203). Although Recasens and Farnetani concluded that loss of /l/ must be production-based rather than perception-based, Browman and Goldstein have suggested that if an anterior lingual gesture occurs in utterance-final silence, then deletion of that gesture might then become a listener-based sound change (Browman and Goldstein, 1995), see also Ohala (1981), i. e., listeners would reinterpret the auditorily covert gesture as a deletion and might fail to produce the gesture at all in their own speech.

The present study uses ultrasound tongue imaging (UTI) to investigate the role of lingual gesture timing in the audible weakening of coda /r/, which has received less attention than /l/ to date. Most studies look at liquid consonant gesture timing in highly-constrained sets of utterances (Sproat and Fujimura, 1993; Scobbie and Pouplier, 2010; Turton, 2014), often involving the flanking of the liquid under study with front high vowels in order to be able to reliably distinguish between tongue gesture movements associated with the consonant and those associated with the vowel. Posterior articulatory tongue gestures for /r/ (i. e., tongue root retraction gestures) tend to have merged with those of preceding non-high vowels and therefore cannot always be reliably identified. As rhoticity strength shows such strong social stratification in the communities under study and because we did not want participants to be aware that the study concerned /r/, we included words with a wide range of (i. e., both high and non-high) prerhotic vowels. This meant that rather than looking at the timing of the anterior (tip or dorsum raising) and posterior (root retraction) gestures that make up coda /r/, we could only reliably identify the anterior gestures. We opted to quantify the timing of the anterior lingual gesture relative to the offset of voicing, or onset of a following labial consonant; two events that have the potential to audibly mask the anterior /r/ gesture.

B. Rhoticity in Central Scotland

In this study we make use of socially-stratified variation in coda /r/ production that is evident in different sociolects in Central Scotland, in order to study the role of tongue gesture timing variation in /r/ weakening, and how it relates to variation in the acoustic characteristics of /r/ and variation in perceived strength of rhoticity.

For several decades, sociolinguistic researchers have noted weakening of coda /r/ in the vernacular English of Central Scotland (Romaine, 1978; Speitel and Johnston, 1983; Macafee, 1983; Stuart-Smith, 2003, 2007b; Jauriberry *et al.*, 2012). These mainly auditory-acoustic studies have shown that strength of rhoticity is socially stratified, with middle-class (MC) speakers preserving, and even strengthening rhoticity, while working-class (WC) speakers are often

weakly rhotic. By weakly rhotic, we mean that minimal pairs such as *bud/bird* /bʌd/bʌɹd/ and *cod/cord* /kɔd/kɔɹd/ can, for the most part, still be differentiated by local speakers, but are not easily differentiated by those unfamiliar with WC Central Scottish English (see Lennon, 2013), /r/ is so weakly produced that *cod* and *cord* can sound equally /r/-less to those not familiar with the Central Scottish accent. Acoustic and auditory analyses identify some prerhotic vowel modifications that are associated with coda /r/ weakening, such as pharyngealisation of the prerhotic vowel [bʌ^hd] “bird,” or lengthening of the prerhotic vowel. Stuart-Smith’s auditory-acoustic study of WC Glaswegian postvocalic /r/ found that if /r/-ful words were derhoticised, formant values and trajectories for /r/-ful and /r/-less minimal pairs were very similar, but that an acoustic correlate of pharyngealised or uvularised derhotic variants appeared to be F3 raising. Stuart-Smith also found instances of /r/-ful words where there were no changes in formant frequencies or amplitudes throughout the vocalic portion of the vowel, e. g., “heart” produced as [hɑ^hʔ]. Finally, it was found that the vocalic portion of derhoticised /r/-ful words was significantly longer than that of /r/-less words (Stuart-Smith, 2007b). Stuart-Smith (2007b) concluded that “acoustic analysis shows few straightforward links with auditory findings” (Stuart-Smith, 2007b, p. 1452) and that “explaining (the derhoticisation) process will need recourse to articulation” (*ibid.*). In an acoustic analysis of /r/-ful and /r/-less minimal pairs in Glaswegian English, Lennon *et al.* (2015) found that WC /-id/ /-ird/ minimal pairs were easy to differentiate due to the presence of a prerhotic offglide: [bid] “bead” versus [biʌd] “beard.” However, formant tracks throughout the vocalic portion of the word for /-ʌd/ /-ʌɹd/ minimal pairs showed a great deal of similarity, with significant variation found only in the duration of the vocalic portion of the word (/r/-ful tokens were longer) and variation in F2 (F2 was found to be lower throughout the vocalic portion for /r/-ful tokens). It would seem therefore that articulatory analysis is required in order to fully understand Central Scottish derhoticisation.

To date, articulatory analysis of coda /r/ in Central Scottish speech communities has already uncovered articulatory variation of which sociolinguistic and dialectological researchers were unaware. There is a strong tendency for coda /r/ tongue shape to be socially-stratified, with WC speakers tending to produce coda /r/ with tongue-tip/front raised variants of /r/, while MC speakers tend to produce /r/ with tongue bunching (Lawson *et al.*, 2011a, 2014a). Bunched /r/ variants have been identified and studied along with retroflex /r/ for decades in American English, see Delattre *et al.* (1968), Lindau (1985), Mielke *et al.* (2010), and Zhou *et al.* (2008). In studies of American English, these /r/ tongue shape variants appear to be used idiosyncratically, or for coarticulatory reasons, e. g., bunched variants occurring before high front vowels. In Central Scottish English, socially-stratified variation between (i) tip/front raised and (ii) bunched coda /r/ variants has been identified in four separate adolescent ultrasound word list speech corpora, collected between 2007 and 2014 (Lawson *et al.*, 2014b, 2014a, 2011a), including the dataset used in the current study (Lawson *et al.*, 2014a). Figure 1 below illustrates tongue shape variation in the current dataset,

with an average 94% of coda /r/ tongue shapes produced by MC speakers classified as bunched, while 94% of coda /r/ tongue shapes produced WC speakers were classified as “tip up” or “front up.”

The social stratification of tongue configurational variation for coda /r/ might explain some of the audible differences in rhotic strength between WC and MC coda /r/, in that bunched /r/ variants might produce a particularly strong audible impression of rhoticity. Delattre and Freeman, for example, found that a bunched tongue shape type 4, “dorsal bunched with dip” (Delattre and Freeman, 1968, p. 41, Fig. 1) produced the “strongest auditory impression” (*ibid.*, p. 64) of rhoticity. Nevertheless, tongue configuration variation does not explain the extreme degree of weakening found in WC Scottish coda /r/. A tip/front raised /r/ tongue shape is *not* associated with weak rhoticity *per se*, and equally rhotic-sounding tip/front-raised and bunched variants of /r/ are found in other varieties of English (see Delattre *et al.*, 1968; Zhang *et al.*, 2003; Twist *et al.*, 2007). In other UTI-acoustic datasets collected in Central Scotland, a delay in the tongue tip/front raising gesture has been observed and reported in WC speech (Lawson *et al.*, 2011a, 2014a), but never studied systematically, or quantified until now. The dataset used in this study therefore allows us to determine whether lingual gesture timing variation underlies the natural weakening of coda /r/, by analyzing covariation between (a) the timing of underlying tongue gestures, (b) percept of rhoticity, and (c) formant patterns in two speaker groups who weaken and strengthen coda /r/, respectively.

We hypothesise that underlying the weakly-rhotic, WC coda /r/ is a delay in the timing of the tip/front raising gesture, to the extent that it is partly or fully masked by silence after voicing offset, or by the frication or closed phase of a following labial consonant. We also hypothesise that variation in lingual gesture timing will correlate with variation in the auditory percept of rhoticity and we will identify the acoustic correlates of tokens with early and delayed anterior /r/ gestures.

II. METHOD

A. Participants

The Western Central Belt audio-UTI corpus was collected in 2012 in the city of Glasgow, Scotland. We recorded young adolescents (12–13 yrs old), as younger speakers tend to produce the best-quality ultrasound images due to their smaller head size, i.e., there is a shorter distance from the ultrasound probe surface to the tongue surface. Recruiting informants from schools also helped socially stratify our corpus, as schools with affluent vs deprived catchment areas were approached to participate.

Sixteen Glaswegian speakers aged 12–13 were recorded for this study. Half were male and half were female, from schools that were geographically close to one another (within 2 miles). Demographic information, presented in Table I, indicates pupils’ different social backgrounds and potential future socio-economic trajectories.

B. Recording scenario and equipment

Informants were recorded with audio and UTI in an IAC sound-attenuated recording booth at the University of Glasgow. All noisemaking equipment such as the ultrasound machine and PC were located outside of the recording booth. Participants were fitted with an Articulate Instruments Ltd. stabilising headset with the ultrasound probe held in place underneath the chin, to eliminate roll and yaw, and minimise pitch movement of the ultrasound probe in relation to the head (Scobbie *et al.*, 2008).

Single word prompts were presented orthographically to participants one at a time on a monitor. Audio recordings were made using a Beyer-Dynamic Opus 55 headworn microphone. Audio recordings were sampled at 22 kHz. A Mindray DP2200 ultrasound machine, set to NTSC video format, created UTI video at a target rate of 29.97 fps. The frame rate of the UTI video was doubled to 59.94 fps by deinterlacing each

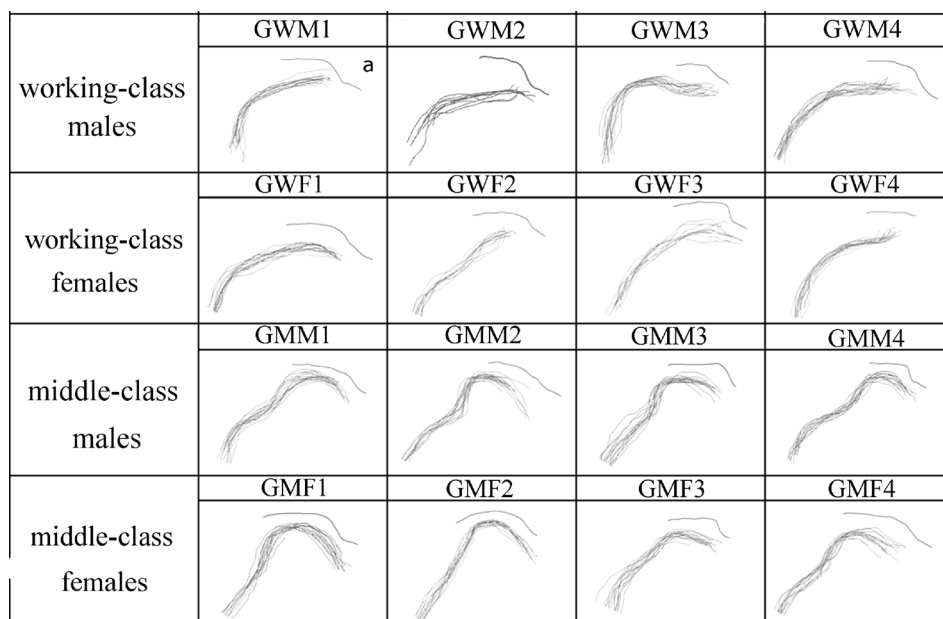


FIG. 1. Bundles of CVr-word midsagittal tongue surface splines, extracted from the tongue surface at the point of maximum constriction for /r/, organized by social class and sex. The uppermost line in each cell represents the midsagittal alveolar ridge and hard palate, with the alveolar ridge to the right of the cell. (a) GWM1 produced only tapped and trilled /r/ variants and was excluded from the study. Figure adapted from Fig. 12-8 (Lawson *et al.*, 2014a).

TABLE I. Key (2011/12) social demographic information pertaining to the two schools involved in the study.

| | WC school | MC school |
|--|------------------|------------------|
| 16 yr old students obtaining 5 or more awards at highest grade | 28% | 79% |
| Students going on to higher education | 22% ^a | 77% ^a |
| Students registered for free school meals ¹ | 30% | 3% |

^aData for 2010/11.

video frame *post hoc*. See Sec. IID for details of audio-video synchronisation.

C. Word list

Our word list had 30 stressed, monosyllabic items containing coda /r/, 25 with a CVr structure, e.g., *bear* and 5 with a CVrC structure, e.g., *herb*, see Table II. There were also 14 /r/-ful nonsense words (not analysed in this study) and 98 distractors. Nine words in the word list contained one of the set of Scottish checked vowels /ɪ, ʌ, ɛ/, which are of particular interest in relation to gesture timing in /r/. The term “checked” refers to the phonotactic specification that vowels do not occur in stressed open word-final syllables, i.e., they must always be followed by a consonant in stressed syllables. Checked vowels tend to be phonetically more lax than unchecked vowels and are also shorter than unchecked vowels in most varieties of English, though not in Scottish English where vowel length is not phonemic (Scobbie *et al.*, 1999). The short, lax phonetic quality of the checked vowels seems to have permitted following /r/ to exert a strong coarticulatory force over them historically. In American English, we see historical coalescence of these vowels with /r/ and merger of the three vowels to [ɚ] (Wells, 1982a, Sec. 6.1.5). Historical changes in Anglo-English relating to the checked vowels are not recoverable, but in present-day Anglo-English, these vowels are merged to [ɜː], known as the NURSE merger (Wells, 1982b, Sec. 3.1.8). In Scottish English, merger/coalescence of /ɪ, ʌ, ɛ/ before /r/ to [ɚ] is a longstanding feature of MC, but not WC speech, and previously thought to be an adaptation toward Anglo-English phonology (see Aitken, 1979). However, an ultrasound-based study by Lawson *et al.* (2013) has shown that coalescence of the checked vowels and /r/ occurs due to the strong coarticulatory force exerted by bunched /r/. Lawson *et al.* (2013) suggested that the timing of the maximum of the rhotic gesture occurs early in the /ɪr, ɛr, ʌr/ syllables where bunched /r/ is used, and not when tip-up/front-up /r/ is used, but this vowel-dependent rhotic timing variation has not been quantified to

TABLE II. Word list items, arranged according to prerhotic vowel.

| Checked prerhotic vowel /ɪr, ʌr, ɛr/ | Unchecked prerhotic vowels | | | | | | |
|---|----------------------------|-------------|-------------|------------|-------------|-------------|-------------|
| | /ɪr/ | /ɛr/ | /ʌr/ | /ɔr/ | /or/ | /ʊr/ | |
| <i>fir</i> | <i>err</i> | <i>ear</i> | <i>air</i> | <i>bar</i> | <i>for</i> | <i>bore</i> | <i>boor</i> |
| <i>firm</i> | <i>her</i> | <i>beer</i> | <i>bear</i> | <i>far</i> | <i>form</i> | <i>more</i> | <i>moor</i> |
| <i>burp</i> | <i>herb</i> | <i>fear</i> | <i>hair</i> | <i>par</i> | <i>or</i> | <i>oar</i> | <i>poor</i> |
| <i>fur</i> | <i>perm</i> | <i>peer</i> | <i>pair</i> | | | <i>pore</i> | |
| <i>purr</i> | | | | | | | |

date. The inclusion of checked and unchecked prerhotic vowel variants in this study, and the inclusion of the fixed factor PRERHOTIC VOWEL with levels checked and unchecked in our statistical analysis, will allow us to quantify the impact of the checked status of the vowel on the gestural timing of coda /r/.

There is some historical evidence that coda /r/ was lost earliest in a preconsonantal position in Anglo-English (Dobson, 1957, Sec. 401). The structure of the word, i.e., CVr or CVrC structure, as in “fir” and “firm,” was therefore included in the statistical analysis of the dataset, as the fixed factor STRUCTURE, in order to identify whether timing of the anterior rhotic gesture is affected by a word’s structure, or whether acoustic and auditory measures are affected by a word’s structure.

We avoided lingual consonants in prompts, to reduce potential coarticulatory effects on /r/.

D. Ultrasound-audio resynchronization

An essential preliminary phase was the resynchronization of the audio and ultrasound video channels and calculation of internal processing lag for the video-based ultrasound machine. The audio-ultrasound recording system, like many others, involves separate audio and video channels, received and processed by a laptop computer. We therefore needed a *post hoc* means of synchronizing audio and video and re-establishing the video frame rate. Both audio and video channels passed through a SynchBrightUp unit (Articulate Instruments Ltd.) which acted as a clapperboard system, superimposing a flash on the video signal and a tone on the audio signal at the beginning of each new recording and encoding information about video frame rate. These signals were later used by Articulate Assistant Advanced ultrasound recording and analysis software (Wrench, 2012) to re-establish the UTI video frame rate and to resynchronise audio and video on each recording by aligning flash and tone. Additionally, video-output ultrasound machines have a variable internal processing delay of several milliseconds, while the data collected at the probe are turned into a video frame.

The mean internal video processing delay of an ultrasound scanner can be estimated using a “tap test”: the microphone capsule was tapped onto the ultrasound probe, and the mean delay from the audio to the visual record of this event was calculated from 100 taps. The DP2200 was found to have an average image processing delay of 20 ms (standard deviation = 14 ms)—a little under one deinterlaced video frame. A -20 ms lag was introduced to the video signal to account for this delay. The variability of the processing delay means that there is a slight random variation in the amount of time it takes the DP2200 to create and output each video frame. Such inconsistencies in synchronization of video and audio act as randomly-distributed noise in the data.

E. Tongue gesture timing annotation

As already mentioned, the present study quantifies the durational difference between the maximum of the anterior rhotic lingual gesture in coda /r/ (tip/front raising or dorsum raising, depending on the type of /r/ involved), and either voicing offset or following consonant onset. In order to measure

this durational difference and normalize the measure, four main temporal events in each coda /r/ token were annotated:

- (1) *rmax*—the visually-determined location of the maximum of the anterior lingual /r/ gesture;
- (2) *V-onset*—the acoustically-determined location of the onset of the vowel in CVr and CVrC words;
- (3) *voice-offset*—the acoustically-determined location of the offset of voicing in CVr words; and
- (4) *C-onset*—the acoustically-determined location of the onset of the closing labial consonant in CVrC words.

These key temporal measurements reveal the relationship between the timing of *rmax* and articulatory events that could render the /r/ gesture partly or completely inaudible (Fig. 2).

V-onset, *voice-offset*, and *C-onset* values were annotated by E.L. using Praat (Boersma and Weenink, 2013), using waveform, spectrogram, and Praat tools such as the pitch and intensity trackers as a guide.

Raw lag duration between *rmax* and an articulatory event that could render the /r/ gesture inaudible was calculated as follows:

CVr words *ear*, *bar*, *pore*, etc. ($\text{lag} = r_{\text{max}} - \text{voice-offset}$)

CVrC words *herb*, *firm*, *form*, etc. ($\text{lag} = r_{\text{max}} - C\text{-onset}$).

Time-normalisation of these raw measurements was also carried out to take account of potentially different speech rates. Normalisation involved dividing the *raw lag* durational measure by the *V-onset* to *voice-offset/C-onset* duration, which means that *normalised lag* is *raw lag* expressed as a proportion of the vowel + /r/ section of the syllable rime. Where *rmax* occurs before *voice-offset* or *C-onset*, as in Fig. 2, left, *raw lag* and *normalised lag* are negative. Where *rmax* occurs after *voice-offset* or *C-onset*, as in Fig. 2, right, *raw lag* and *normalised lag* are positive.

F. Auditory analysis

In order to confirm the hypothesis that delayed anterior /r/ gestures are associated with weakened audible rhoticity,

all tokens of /r/ were rated on a rhoticity index, using a Praat *multiple forced choice* (MFC) experiment interface. The interface presented randomised anonymous tokens to three classifiers (the authors), two speakers of Standard Scottish English from the Central Scotland, and one speaker of Standard Southern British English with 20 years' experience studying Central Scottish speech. Each token was rated on a rhoticity index –5-point ordinal scale, arranged from /r/-less to strongly /r/-ful. We used an r-index of ordered categories ranging from least to most /r/-ful to obtain greater consistency between raters and hence gain a more meaningful capture of their ratings. The r-index itself is based on gradient r-indexes in Scottish rhoticity literature, e.g., Stuart-Smith (2007b) and Jauriberry *et al.* (2012). Category labels registered in the MFC as numerical codes ranging from 1 (no /r/) to 5 (schwar). Despite tapped and trilled variants being stereotypically associated with Scottish speech, only one speaker, GWM1, produced them, but never in his spontaneous speech. He was also the only speaker who identified /r/ as the focus of the study, therefore this speaker was removed from the study.

In the rhoticity index, henceforth *r-index*, *no /r/* referred to no auditory percept of /r/, i.e., the word sounded as if it ended in a (non-rhoticised) vowel, e.g., [fʌ] “fur,” [fiə] “fear”; *derhoticised* referred to variants where there was an audible hint of /r/, or some other feature that could be associated with rhoticity in Glaswegian English, such as pharyngealisation or velarisation of the prerhotic vowel, but no clear rhotic segment, e.g., [fʌʕ], [fiʕ];² *alveolar* referred to a (post)alveolar approximant with a less strong rhotic quality, e.g., [fʌɹ], [fiɹ] than the *retroflex* approximant, which referred to a variant that sounded like a strongly rhotic approximant, e.g., [fʌɻ], [fiɻ], and, finally, *schwar* was a central rhotic vowel [ɤ] in place of a Vr sequence, e.g., [fɤ:], [fiɤ:]. From a phonological point of view, [ɤ] might be considered by some to be a vocalisation of /r/ and therefore weaker than a retroflex approximant; however, in this study [ɤ] was considered by the raters to have very strong audible rhoticity. This phonetic variant is particularly associated with an underlying bunched articulation of /r/, occurring after the checked vowels /ɪ, ɛ, ʌ/ in Scottish English

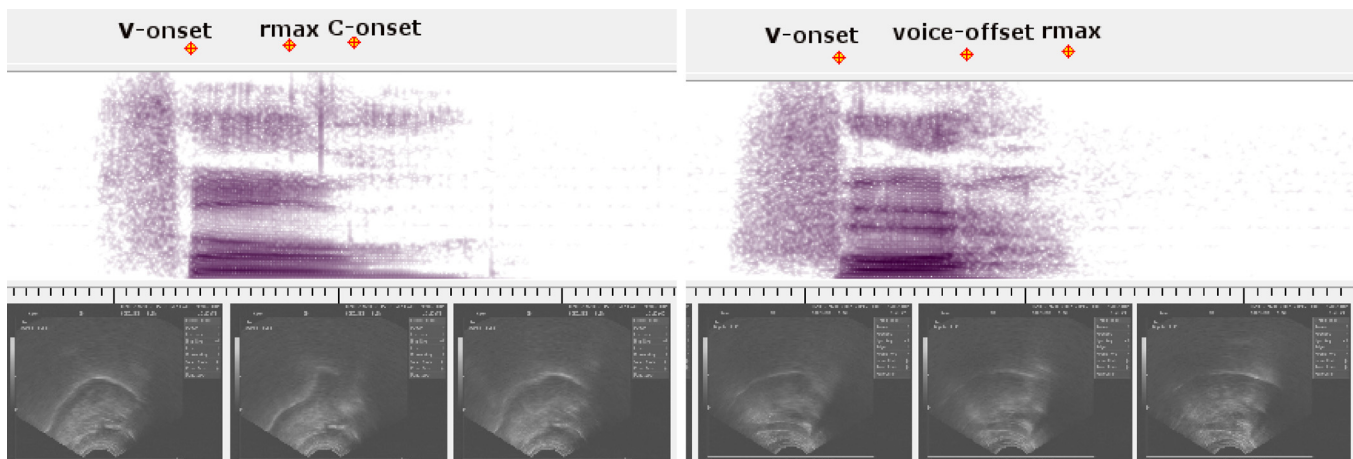


FIG. 2. (Color online) (Left) MC speaker (GMM2) “firm” [fɜ:m] showing V-onset, rmax, and C-onset annotations with spectrogram. (Right) WC speaker (GWF4) “fir” [fɪʕ] showing V-onset, rmax, and voice-offset annotations with spectrogram. Tickmarks below show the alignment of the UTI keyframes with the acoustic signal, below are sample UTI images showing the progression of the articulations through the recording.

(see Lawson *et al.*, 2013). As [ɚ] represents a coalescence of the vowel and /r/, it is highly likely to involve an early anterior lingual /r/ gesture, close to the syllable nucleus.

(C)Vr(C) word tokens were classified by each auditory rater over several sessions and a mean rhoticity index value was calculated for each token. Inter-rater reliability was gauged using Krippendorff's alpha (Hayes and Krippendorff, 2007) using the *irr* package (Gamer *et al.*, 2012) in R (R Core Team, 2013), showing a moderate level of inter-rater reliability between the three raters $\alpha=0.754$. As the mean *r-index* scores are ranked ordinal data, rather than interval data, we did not include *r-index* in the mixed effects modeling; however, it is included in a nonparametric correlational analysis, Sec. III A. We carried out nonparametric Kruskal-Wallis tests on *r-index* with CLASS, SEX, PRECEDING VOWEL, and STRUCTURE as fixed factors, see Sec. III D.

G. Acoustic analysis

Wherever the articulatorily-determined *rmax* occurred before the offset of voicing, or onset of a following consonant, formants one to five were measured by hand using Praat at the same temporal point of the articulatory *rmax* annotation. Measurement of the first five formants involved inspection of the spectrogram alongside narrowband fast Fourier transform (FFT) and Linear Predictive Coding (LPC) spectra. Close inspection of spectra was required, as, for many strongly /r/-ful variants, *F2* and *F3* were not easy to differentiate using only a spectrographic representation, due to their proximity. Where *rmax* occurred after *voice-offset/C-onset*, formants one to five were measured just before *voice-offset/C-onset*. The first five formants were measured, rather than just the traditional first three formants for rhotics, as an MRI-based study by Zhou *et al.* (2008), suggested that retroflex and bunched variants may be acoustically distinct and auditorily discriminable due to variation in the higher formants, specifically *F4* and *F5*, with bunched /r/ being characterised by a greater acoustic distance between *F3* and *F4*, and a lesser acoustic distance between *F4* and *F5*, than retroflex /r/. Given the socially-stratified bunched-retroflex variation in our dataset (Lawson *et al.*, 2014a), by measuring *F4* and *F5* in addition to *F1–F3* we aimed to take account of the potential impact of tongue shape variation on the acoustic signal and to determine whether tongue shape was also having an impact on perceived strength of rhoticity. In addition, we hoped to contribute to the understanding of the acoustics of bunched-retroflex variation in /r/. However, we encountered difficulties in measuring *F5*, particularly at voicing offset and no measurement could be taken for around 10% of tokens, mainly tokens from WC speakers. We were not completely confident of the accuracy of *F5* measures and initial linear effects modeling showed a lack of significant findings for *F5* for both fixed and random factors, therefore we decided to analyse *F1–F4* only.

Whilst differences in vocal tract size between speakers are likely to result in small differences in formant values for /r/ for this early adolescent speaker group, just as they would for vowels (e.g., Adank *et al.*, 2004), there is currently no accepted method of acoustic normalisation for rhotics,

particularly the higher formants. In our mixed effects regression analysis (see Sec. II H) we included the fixed effect of SEX with two levels: male and female, and we also included SPEAKER as a random factor. Arguably, given that the speakers were aged 12–13, the inclusion of the random factor SPEAKER may be more important than the fixed factor SEX, as in this early stage of adolescence, sex-based vocal tract differences are less predictable. Anecdotally, the authors often misidentified males as females and vice versa in the study when listening to audio recordings.

H. Statistical analysis

In order to examine the relationship between the articulatory, auditory, and acoustic variables in the study, and to test the hypothesis that gesture delay is responsible for coda /r/ weakening, we first carried out a Spearman's correlational analysis of all the dependent measures with Bonferroni corrections to take account of the fact that a series of multiple tests were run simultaneously.

Thereafter, given that coda /r/ weakening and strengthening are particularly associated with WC and MC speech, respectively, we used mixed effects modeling to determine whether social class was a significant predictor of variation across the articulatory and acoustic measures taken. In these models, we also took into account other features of our dataset that might potentially have a significant effect on the dependent measures. Our six dependent variables were: *raw lag*, *normalised lag*, and *F1 to F4*. The following four fixed factors were included: (1) PRECEDING VOWEL (checked /ɪ/, /ɛ/, /ʌ/, or unchecked /i/, /æ/, /a/, /o/, /ɔ/, /e/), (2) STRUCTURE of the word [CVr, e.g., *moor* or (C)VrC, e.g., *form*], (3) CLASS (WC or MC), and (4) SEX (male or female). We also tested for interactions between CLASS and SEX in order to identify whether males and females were behaving differently within their class groups, as has been suggested (Jauriberry *et al.*, 2012; Lawson *et al.*, 2014a), and between CLASS and PRECEDING VOWEL, because we expected vowel-to-/r/ coalescence where /r/ follows checked vowels in the MC group, which we would expect to have an impact on all of the dependent variables. A variance-inflation-factors test was carried out in order to check for collinearity among fixed factors, where any factor obtaining a value >2 would be removed. All fixed factors were kept.

Mixed effects modeling takes account of aspects of the experimental design that involve sampling within a population, allowing the inclusion of random factors that are not generalizable to the wider population in the way that fixed factors such as CLASS, SEX, etc., would be. SPEAKER and PROMPT were included as random factors to prevent extremes of variation in the behavior of particular speakers or extremes of variation in the production of particular prompts having an undue effect on statistical results, see Drager and Hay (2012) and Hay (2011). The inclusion of SPEAKER as a random factor was particularly important in the construction of models for the formant data, as it helped to remove formant variation that might be attributable to variation in vocal-tract length. SPEAKER was found to be significant for all factors. PROMPT was significant for *F1 – F3* only. We

TABLE III. Correlation matrix for WC speakers, showing Spearman's r_s . *** = significant to the $p < 0.001$ level, * = significant to the $p < 0.05$ level. Bold font indicates correlations where $-0.5 \geq r_s \geq 0.5$.

| raw | norm | r-index | F1 | F2 | F3 | F4 |
|---------|----------------|-----------------|-----------------|-----------------|-----------------|----------|
| raw | 0.98*** | -0.69*** | 0.21 | -0.50*** | 0.57*** | -0.30*** |
| norm | | -0.69*** | 0.18 | -0.48*** | 0.57*** | -0.28*** |
| r-index | | | -0.25*** | 0.44*** | -0.65*** | 0.30*** |
| F1 | | | | -0.08 | 0.25*** | -0.14* |
| F2 | | | | | -0.06 | 0.09 |
| F3 | | | | | | -0.31*** |
| F4 | | | | | | |

used the lme4 package in R (v3.1.2) followed by the step() function to find the models that best fit the data. The auditory measure, the mean *r-index* scores, constitute ordinal rather than interval data. We carried out nonparametric Kruskal-Wallis tests on *r-index* with CLASS, SEX, PRECEDING VOWEL, and STRUCTURE as fixed factors.

We will only report on effects and correlations that were found to be significant.

III. RESULTS

A. Correlational analysis

Most dependent variables were found to correlate with one another, see Table III.

We inspected and interpreted only the strong correlations, which we define here as those which were not only significant, but also showed $-0.5 \geq r_s \geq 0.5$. These strong correlations are displayed in the scatterplots in Fig. 3.

The two articulatory measures, *raw* and *normalised lag*, were unsurprisingly the most closely correlated $r_s = 0.98$, $p < 0.001$. After this, the correlation matrix (Table III) and plots (Fig. 3) reveal a correlational triangle between articulatory and auditory variables and the acoustic variable most strongly associated with rhoticity, *F3*, and to a lesser extent, *F2*. There were four key findings from the correlational analysis: (1) There was a strong negative correlation between *r-index* and *lag* ($r_s = -0.69$, $p < 0.001$), see Fig. 3(d); the greater the *lag*, the less /r/-ful sounding the word. (2) There was a strong negative correlation between *F3* and *r-index* ($r_s = -0.65$, $p < 0.001$), see Fig. 3(c); the higher the *F3*, the less /r/-ful sounding the word. (3) There was a strong positive correlation between *lag* and *F3* ($r_s = 0.57$, $p < 0.001$), see Fig. 3(a); the greater the *lag*, the higher the *F3*. (4) There was a strong negative correlation between *F2* and *raw lag* ($r_s = -0.50$, $p < 0.001$), see Fig. 3(b); the greater the *raw lag*, the lower the *F2*. In summary, our hypothesis was correct; the greater the *lag*, the higher *F3* and the less /r/-ful sounding the word.

The correlation plots (Fig. 3) also illustrate social stratification of the dependent variables *raw lag*, *r-index*, *F2*, and *F3*, as the WC speakers' tokens, black circles, are shown to

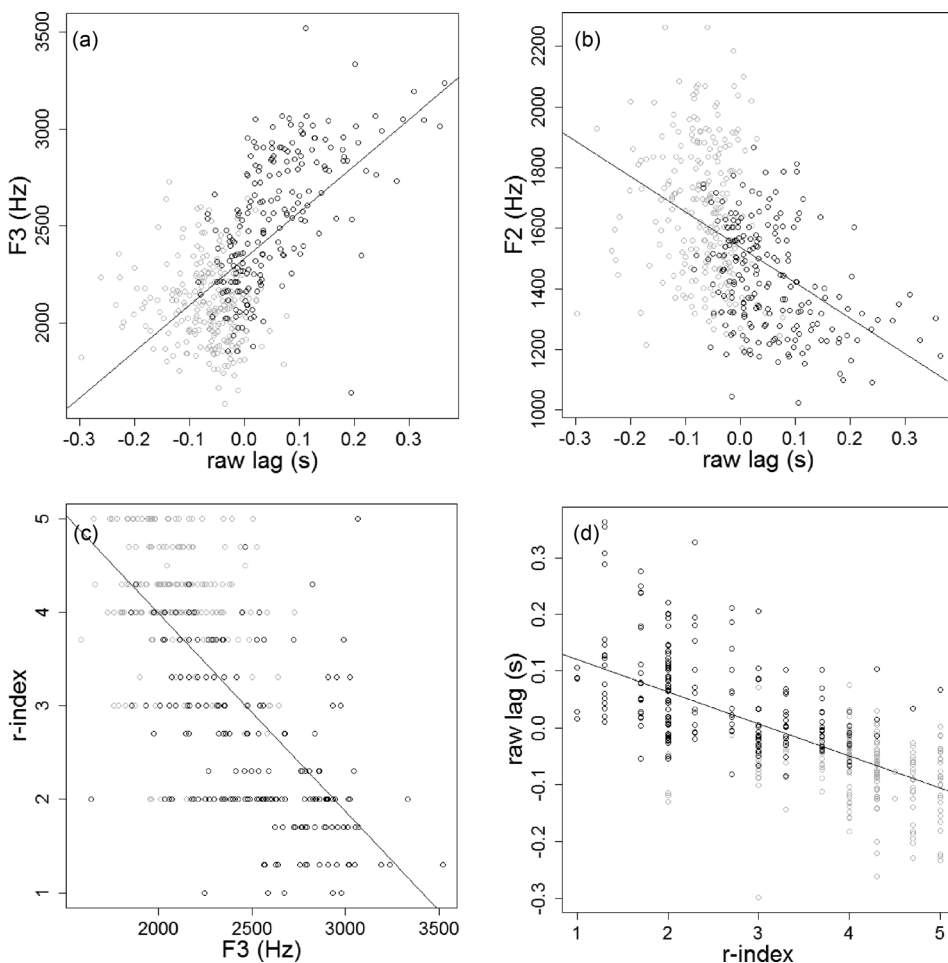


FIG. 3. Scatterplots of strongest correlations in WC dataset: (a) raw lag and *F3*, (b) raw lag and *F2*, (c) *F3* and *r-index*, and (d) *r-index* and raw lag. WC data points are black, while MC data points are gray.

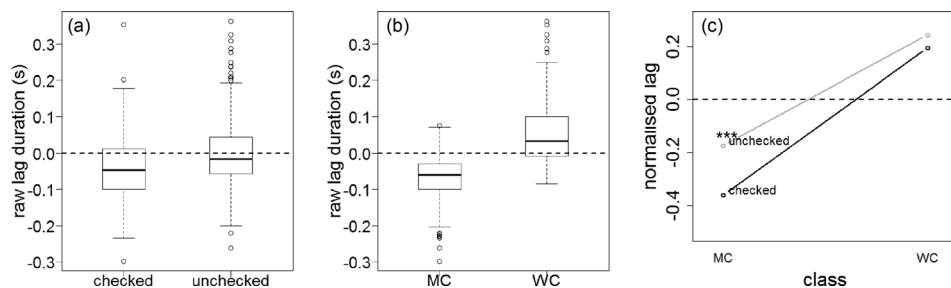


FIG. 4. Boxplots of (a) raw lag by PRECEDING VOWEL, (b) raw lag by CLASS, and (c) interaction plot for CLASS and PRECEDING VOWEL, showing adjusted means for normalised lag. Checked = /t,ɛ,ʌ/, unchecked = all other vowels; MC = middle class, WC = working class. *** = significant to the <0.001 level. N = 428.

have shorter lags, lower *r-index* scores, lower *F2*s and higher *F3*s than the MC speakers' tokens, gray circles. The following sections confirm the impact of social class on these measures, also in interaction with linguistic factors.

Other significant, but weaker, correlations were found where $-0.5 < r_s > 0.50$. *F2* was positively correlated with *r-index* ($r_s = 0.44, p < 0.001$); the higher *F2*, the more /r/-ful sounding the word. *F3* was negatively correlated with *F4* ($r_s = -0.31, p < 0.001$); the higher *F3*, the lower *F4*. *Raw lag* was negatively correlated with *F4* ($r_s = -0.30, p < 0.001$); the greater the lag, the lower *F4*, and finally, *F4* was positively correlated with *r-index* ($r_s = 0.30, p < 0.001$); the higher *F4*, the more /r/-ful sounding the word.

B. Articulatory: Gesture timing—raw and normalised lag

Figure 4 presents as boxplots the distribution of the *raw lag* for the significant factors: PRECEDING VOWEL ($F = 42.40, p < 0.001$, checked vowel estimated mean *raw lag* = $-0.034 \text{ s} \pm 0.013 \text{ s}$, unchecked vowel = $-0.0019 \text{ s} \pm 0.0128 \text{ s}$) and CLASS ($F = 22.98, p < 0.001$, WC estimated mean *raw lag* = $0.045 \text{ s} \pm 0.019 \text{ s}$, MC = $-0.077 \text{ s} \pm 0.017 \text{ s}$). For *normalised lag*, a significant interaction was found for PRECEDING VOWEL and CLASS ($F = 11.12, p < 0.001$). *t*-tests showed significance for checked and unchecked vowels in MC speech only $t = 6.62, p < 0.001$.

In Fig. 4 above, where data points occur below the broken horizontal lines, the maximum of the /r/ gesture occurred before voicing offset in CVr words, or before the onset of the final consonant in (C)VrC words. Where data points occur above this line, some or all of the anterior lingual /r/ gesture occurred after voicing offset, or during the articulation of a following labial consonant. In the former case, we would expect /r/ tokens to be audible, in the latter case, we would expect the /r/ to be either completely inaudible, in cases where the maximum of the /r/ gesture is very delayed, or, for the /r/ to be audibly weakened to varying extents, depending on how much of the anterior lingual gesture is masked.

C. Acoustics: Formant measures

1. F1

Figure 5 below shows a significant interaction for *F1* between CLASS and PRECEDING VOWEL $F = 22.11, p = 0$. MC speakers showed a lower estimated mean *F1* (562 Hz, $\pm 33 \text{ Hz}$) than WC speakers (667 Hz, $\pm 35 \text{ Hz}$). /r/ after checked vowels also had a significantly higher *F1* than after unchecked vowels, $t = 3.56, p < 0.001$ for MC speakers only.

2. F2

Figure 6 shows as boxplots the distribution of *F2*s for the significant factors: STRUCTURE ($F = 11.6, p < 0.01$, CVr estimated mean *F2* = $1550 \text{ Hz} \pm 37 \text{ Hz}$, CVrC = $1445 \text{ Hz} \pm 43 \text{ Hz}$), and SEX ($F = 6.88, p < 0.05$, female estimated mean *F2* = $1588 \text{ Hz} \pm 48 \text{ Hz}$, male = $1408 \pm 52 \text{ Hz}$). A significant interaction for CLASS and PRECEDING VOWEL was identified ($F = 9.43, p < 0.01$), see Fig. 6(c), but *t*-tests showed no significant differences between /r/s after checked and unchecked vowels in MC and WC tokens.

3. F3

Figure 7 below shows as boxplots the distribution of *F3* for the significant factors: STRUCTURE ($F = 22.75, p = 0$, CVr estimated mean *F3* = $2320 \text{ Hz} \pm 48 \text{ Hz}$, CVrC = $2175 \text{ Hz} \pm 52 \text{ Hz}$) and SEX ($F = 5.27, p < 0.05$, female estimated mean *F3* = $2353 \text{ Hz} \pm 64 \text{ Hz}$, male = $2142 \pm 69 \text{ Hz}$). There was a significant interaction for CLASS and PRECEDING VOWEL ($F = 7.61, p < 0.01$), see Fig. 7(c). A *t*-test showed a significant difference in *F3* between checked and unchecked vowels for the WC speakers only $t = 3.62, p < 0.001$.

4. F4

Figure 8(a) below shows as a boxplot the distribution of *F4*s for the significant factor: STRUCTURE ($F = 4.46, p < 0.05$, CVr estimated mean *F4* = $3845 \text{ Hz} \pm 63 \text{ Hz}$, CVrC = $3757 \text{ Hz} \pm 70 \text{ Hz}$). There was also a significant interaction

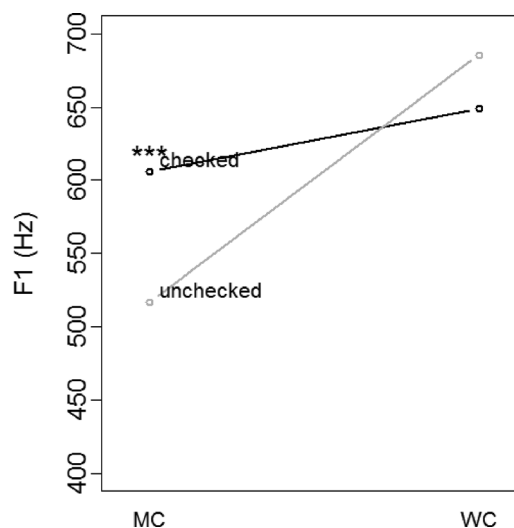


FIG. 5. Interaction plot for *F1* of CLASS and PRECEDING VOWEL. N = 430.

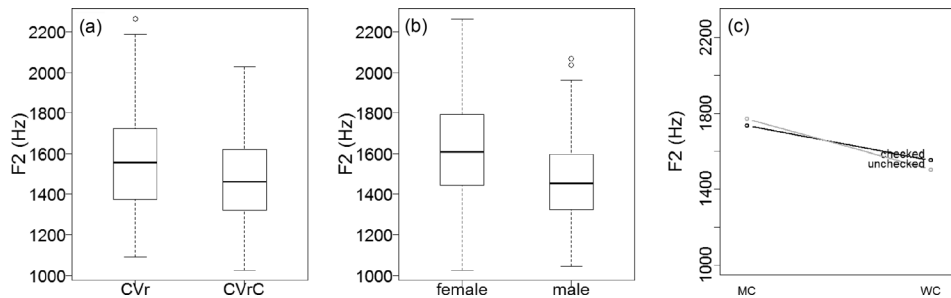


FIG. 6. Boxplots of mean F_2 by (a) STRUCTURE, (b) SEX, and (c) interaction plot for CLASS and PRECEDING VOWEL, showing adjusted means for F_2 . $N = 429$.

between CLASS and PRECEDING VOWEL ($F = 4.7$, $p < 0.05$), with a significant difference between checked and unchecked vowels in the WC group $t = 3.33$, $p = 0.001$, Fig. 8(b).

D. Auditory measures: r -index score—Kruskal-Wallis

We were unable to include the ordinal r -index data in the mixed effects modeling; however, we carried out a series of Kruskal-Wallis nonparametric tests to explore the effect of each of our fixed factors on r -index. Figure 9 below shows as boxplots the distribution of the r -index score for the significant factors: CLASS ($\chi^2 = 217.24$, $p < 0.001$), with WC speakers heard as less /r/-ful than MC speakers; SEX ($\chi^2 = 7.68$, $p < 0.01$), with females heard as less /r/-ful than males, and PRECEDING VOWEL ($\chi^2 = 5.94$, $p < 0.05$), with tokens containing checked vowels heard as more /r/-ful than those containing unchecked vowels.

IV. SUMMARY AND DISCUSSION

Previous articulatory research suggested two potential mechanisms underlying the audible weakening of coda liquids: gesture reduction (Browman and Goldstein, 1995) and gesture delay (Recasens and Farnetani, 1994; Krakow, 1989; Delattre and Freeman, 1968). Observations from a decade of articulatory study of postvocalic /r/ in Central Scotland, where different sociolects have markedly different strengths of rhoticity, suggested that gesture delay could be a key articulatory mechanism underlying coda /r/ weakening (Lawson et al., 2014a). Additionally, variation in gesture timing affecting coda /r/ has been observed in other languages, such as Dutch (Scobbie et al., 2009). We also know that studies of coda /l/ in English have found apical gestures for /l/ to be so delayed in some utterance-final positions that there is no evidence of them in the acoustic signal (Recasens and Farnetani, 1994). Previous articulatory studies of coda /r/ in WC Central Scotland suggested that variation in timing might also be responsible for its

weakening (Lawson et al., 2014a); however, timing variation in Scottish coda /r/ has not been quantified until now.

We hypothesised that underlying the weakly-rhotic WC coda /r/s in our dataset, there would be a delay in the timing of the anterior lingual gesture, and we hypothesised that this variation in lingual gesture timing would correlate with variation in auditory percept of rhoticity. We also investigated the spectral characteristics of /r/ and how they correlated with gesture delay and percept of rhoticity. In the following discussion, we report on correlations between the articulatory, auditory, and acoustic variables in our study and then we consider our findings in terms of our independent social and linguistic variables.

A. Correlation test findings: Articulatory-auditory-acoustic relations for /r/

Correlation tests revealed the articulatory-acoustic-auditory relationship that our study aimed to uncover. Significant strong negative correlations were found between raw/normalised lag and r -index score, between r -index score and F_3 , and there was a significant strong positive correlation between raw/normalized lag and F_3 . There was also a strong negative correlation between raw lag and F_2 . In other words, the longer the lag, the higher F_3 , the lower the F_2 and the lower the r -index score (less /r/-ful sounding the word). Correlation plots also indicated that articulatory lag, acoustic F_2 and F_3 , and auditory r -index were socially stratified, with WC speakers' data points showing longer lags, lower F_2 s, higher F_3 s, and lower r -index scores than the MC speakers.

These findings confirm the importance of F_3 , and to a lesser extent, F_2 , in the perception of rhoticity. Tokens such as WC speaker GWF1's "fir" [fɪʳ], shown in Fig. 2, illustrates the impact of anterior lingual gesture delay on F_2 and F_3 in the acoustic signal. This token was unanimously rated 2 on the rhoticity index (i.e., "derhoticised") by the authors. Close to offset of voicing, F_3 is high and F_2 and F_3 are far apart, although there is a very slight lowering of F_3 and

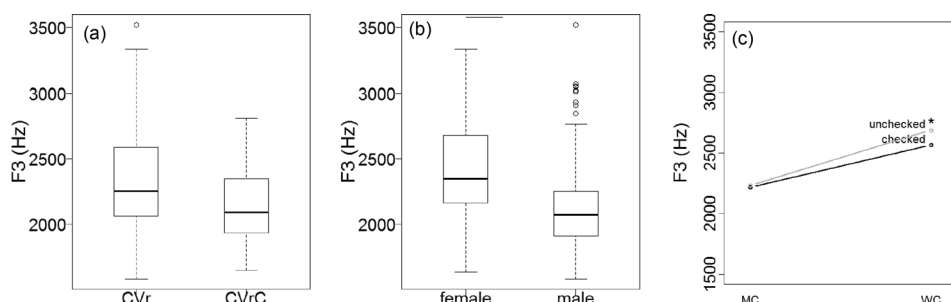


FIG. 7. Boxplots of mean F_3 by (a) STRUCTURE, (b) SEX, and (c) interaction plot for CLASS and PRECEDING VOWEL, showing adjusted means for F_3 . $N = 428$.

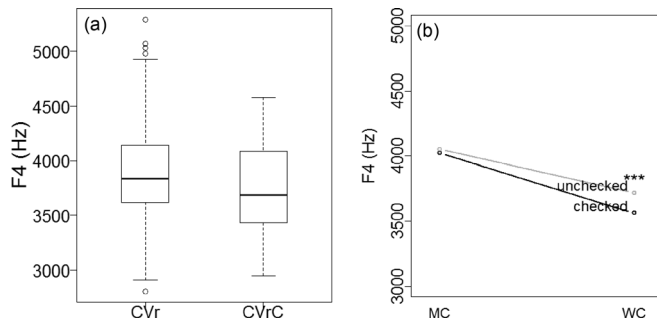


FIG. 8. Boxplots of F_4 by (a) STRUCTURE and (b) interaction plot for CLASS and PRECEDING VOWEL, showing adjusted means for F_4 . $N = 430$.

raising of F_2 just before voicing offset. Ongoing formant changes (continued rising F_2 and falling F_3) are visible on the spectrogram thanks to a breathy exhale following voicing offset. This breathiness is not a systematic feature of WC pronunciation, and therefore not likely to be a general means by which delayed /r/ articulations are cued, but it shows that articulatory movements associated with F_3 lowering and F_2 raising continue after voicing has ceased. For many of the WC tokens, only the initial stages of F_3 lowering and F_2 raising were present before voicing offset, or final-consonant onset. In extreme cases, all F_3 lowering and F_2 raising is masked by voicing offset or by a final labial consonant. This example in Fig. 2, and the findings of this study in general, help illustrate why trained phoneticians have coded tokens of /r/ words in such different ways in auditory-only analysis of Central Scottish speech (e.g., [Stuart-Smith, 2007b](#)).

The finding that anterior lingual gesture delay results in gradient truncation of the acoustic cues associated with rhoticity might be interpreted by some in terms of a production-perception loop, whereby the speaker delays the anterior lingual gesture, part or all of the gesture becomes inaudible, the listener fails to perceive the covert /r/ and begins to produce /r/-ful words without /r/, i.e., the listener as the source of sound change ([Ohala, 1981](#)). Our research and the work of others suggest that such listener reinterpretation is not the main driver of sound change in this community. Mimicry studies carried out by the authors found that very weakly /r/-ful audio tokens were rarely mimicked as /r/-less ([Lawson et al., 2011b](#)). Top-down information, such as lexical access, is likely playing a part in /r/ preservation, as the only weakly /r/-ful token that was mimicked as /r/-less in (*ibid.*) was the only word-list stimulus that had a meaningful minimal-pair counterpart, i.e., “hurt” mimicked as “hut.” However, in a subsequent mimicry experiment involving mimicry of nonsense words with weak /r/s and covert gestures, /r/-less

mimicry was also rare ([Lawson et al., 2014b](#)). [Lennon \(2013\)](#) has shown, furthermore, that those within the Central Scotland speech community can reliably distinguish between /r/-ful and /r/-less minimal pairs, while speakers from outside of the community perform at chance level. This set of findings from mimicry and perceptual experiments perhaps go some way toward answering the perennial question of why sound change does not happen more often. It would seem that lexical access can prevent the reanalysis of /r/-ful to /r/-less when changes to gesture timing weaken the acoustic features associated with rhoticity, but also that even minimal acoustic changes in F_2 and F_3 , or variation in prerhotic vowel duration, can cue rhoticity for some listeners.

Significant correlations relating to F_4 potentially contribute to our understanding of the impact of rhotic tongue shape on acoustics. These correlations suggest that the MC speakers in the study, who predominantly use bunched rhotic variants, see Sec. 1B and Fig. 1, produce rhotics with the highest F_4 values. It was found that F_3 and F_4 had a significant negative correlation, that F_4 was significantly negatively correlated with lag and that F_4 was positively correlated with r -index. In other words, the lower F_3 , the higher F_4 , the shorter the lag the higher F_4 , and the higher F_4 , the higher the r -index score (more /r/-ful the word). These correlational findings tend to support [Zhou et al. \(2008\)](#) regarding their conclusions concerning the acoustic characteristics of bunched /r/, namely that there is a greater distance between F_3 and F_4 for bunched /r/ than for retroflex /r/.

B. Social and linguistic factors in coda /r/ weakening

Speaker social class is rarely considered in phonetic, and especially lab-based, articulatory studies of consonant weakening. In fact, articulatory phonetic accounts of speech sounds are often based on analysis of “standard” or prestigious varieties of languages and on non-stratified convenience samples. Changes in the production of coda /r/ in two Central Scottish sociolects provided an ideal opportunity to assess the impact of lingual gesture timing variation in coda /r/ weakening and strengthening. Earlier articulatory studies had suggested that, in addition to social stratification of coda /r/ tongue shape, there might also be underlying variation in the timing of the anterior lingual gesture for coda /r/, with WC speakers, in particular, delaying the anterior gesture ([Lawson et al., 2014a](#)). Due to the fact that tongue shape is socially-stratified in our dataset, see Fig. 1, we were also able to gather some evidence regarding the impact of tongue shape on formant structure of /r/.

Results confirmed our hypotheses regarding the effect of CLASS on the dependent variables studied; CLASS was

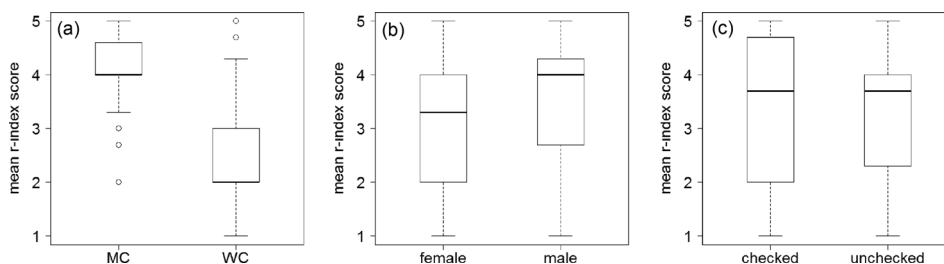


FIG. 9. Boxplots showing mean r -index scores by (a) CLASS, (b) SEX, and (c) PRECEDING VOWEL. r -index category 5 = strongly /r/-ful, category 1 = no audible rhoticity. $N = 408$.

found to have a significant effect for *raw* and *normalised lag* and $F1$ – $F4$. Nonparametric tests on the *r-index* dependent variable also showed that CLASS was a highly significant predictor of audible strength of /r/. WC speakers showed significantly greater gesture *lag*, lower *r-index* score, higher $F1$, lower $F2$, higher $F3$, and lower $F4$ s than MC speakers. The acoustic characteristics of the MC dataset are consistent with the findings of Heselwood (2009) regarding rhotic perceptual cues and auditory integration (Bladon, 1983) among formants 1–3. Heselwood found that perception of rhoticity depends on $F2$ being distant enough from $F1$ to avoid auditory integration, while $F2$ and $F3$ must be close enough for auditory integration to take place, resulting in a perceptual peak in the $F2$ region. Additionally, as mentioned in Sec. IV A, a study by Zhou *et al.* (2008) suggests that a lower $F4$, closer to $F3$, is associated with a retroflex (here “tip/front raised”), rather than a bunched articulation for /r/. Zhou *et al.* obtained MRI and acoustic tokens of retroflex and bunched /r/, respectively, from two adult American males. Spectral analysis showed that for a bunched /r/, there was a much greater distance between $F3$ and $F4$, and a much narrower distance between $F4$ and $F5$, compared to the retroflex acoustics. Unfortunately, variation in $F5$ was not reported in our study, as we were not confident of the accuracy of $F5$ measures, particularly in the WC dataset where measures were often taken at voicing offset.

Social stratification of bunched and tip-up /r/ in our recordings might also explain significant interactions between the factors CLASS and PRECEDING VOWEL for normalised lag. Previous research of rhoticity in the adolescent Central Scottish speech community showed that bunched /r/ exerts a strong coarticulatory pressure on preceding checked vowels, resulting in both coalescence of the vowel and the following /r/, and the neutralisation of a three-way prerhotic contrast—/r, ɛr, ʌr/ to [ɤ] (Lawson *et al.*, 2013). This vowel +/r/ coalescence causes the point of maximum constriction of the anterior lingual gesture in /r/ to occur close to the nucleus of the syllable. We have even observed the maximum of the anterior /r/ gesture occurring as early as during aspiration in a MC pronunciation of the word *purr* [p^hɤ] in another Central Scottish UTI dataset. It was therefore unsurprising that the anterior lingual gesture for /r/ occurred significantly earlier in words containing checked vowels than in those containing unchecked vowels for MC speakers, but not for WC speakers.

We might expect the impact of SEX, assuming smaller vocal tract cavities in girls than boys, to be reflected in an overall raising of the acoustic measures for females. We did find significantly higher values for females for $F2$ and $F3$, which is consistent with this assumption, but $F1$ and $F4$ showed no SEX effect. To some extent, this is likely to be due to the fact that our speakers were in early adolescence, at an age where females often temporarily outgrow their male peers, and we observed a lot of variation in height in the male and female cohorts. Listeners also often misidentified speaker gender when listening to the audio recordings. It is also possible that differences in underlying tongue shape for /r/ were having a bigger impact on vocal tract resonances

than vocal tract length, see Stuart-Smith (2007a) for another example of this phenomenon.

STRUCTURE was not found to affect *raw* or *normalised lag*, or *r-index*. The effect of STRUCTURE on acoustic variables was not straightforward, with the acoustic and perceptual effects of STRUCTURE appearing to partly contradict one another; significantly lower $F2$ s were found in CVrC words, which fit in with weaker rhoticity in these words, but the significantly lower $F3$ s in CVrC words, we would expect to be associated with stronger rhoticity. One explanation for this apparent contradiction between acoustic and auditory findings might be that, as raters heard entire word tokens, they might have had better auditory perception of delayed /r/ in CVr words resulting from formants audible after voicing offset, if noise was present, e.g., breathy exhalation, see Fig. 2, right. This kind of information would not be available for CVrC words with the complete masking effect of a final consonant.

V. CONCLUSION

This study of the phonetic basis of coda lenition identifies gesture delay as a key mechanism for coda /r/ weakening, affecting primarily the third and second formants by causing their maximum and minimum values, respectively, to be masked by other speech events such as voicing offset, or onset of a following consonant. Furthermore, statistical evidence concerning acoustic variables in this study, taken alongside a tongue-shape analysis of the same data (Lawson *et al.*, 2014a), supports the findings of Zhou *et al.* (2008) that underlying bunched tongue configuration is reflected in a greater separation of $F3$ and $F4$, than for tip/front-raised /r/ variants.

Our study presents a picture of /r/ weakening through change in gesture timing that causes tongue gestures in WC speech to be partially auditorily masked, or even auditorily covert. It might be tempting to assume that further weakening will occur through a perception-production loop where perceptual reinterpretation of covert articulations occur, but mimicry studies carried out by the authors (Lawson *et al.*, 2011b, 2014b) and perceptual studies, Lennon (2013), suggest that this is not happening and that speakers from these communities do not, on the whole, reinterpret weakly /r/-ful words as /r/-less. The results of the statistical tests in this study consistently showed CLASS to be the dominant predictor of variation in /r/ gesture timing, formant variation, and audible rhotic strength; in other words, it seems that speakers are exploiting lingual gesture timing, as well as using tongue shape, to index social information. We therefore conclude that social factors are the main driving force behind timing variation in rhotics in this speech community.

ACKNOWLEDGMENTS

This work was supported by Economic and Social Research Council Grant Nos. RES-062-23-3246 and ES-N008189-1.

¹Qualification for free school meals is a widely-used indicator of socioeconomic deprivation, as it requires parents/guardians of students to be in receipt of income support, see Taylor (2017).

²N.B. pharyngealisation does not always have to be present for the derhoticised category, but often is, see Speitel and Johnston (1983). The derhoticised category refers to an audio token where there is some hint of an /r/ present, without there being a clearly audible /r/ segment.

- Adank, P., Smits, R., and Van Hout, R. (2004). "A comparison of vowel normalization procedures for language variation research," *J. Acoust. Soc. Am.* **116**, 3099–3207.
- Aitken, A. J. (1979). "Scottish Speech: A historical view with special reference to the Standard English of Scotland," in *Languages of Scotland*, edited by A. J. Aitken and T. McArthur (W&R Chambers Ltd., Edinburgh), pp. 85–118.
- Bladon, A. (1983). "Two-formant models of vowel perception: Shortcomings and enhancements," *Speech Commun.* **2**, 305–313.
- Boersma, P., and Weenink, D. (2013). "Praat: Doing phonetics by computer" [Computer program]. Version 6.0.23. <http://www.praat.org/>.
- Browman, C. P., and Goldstein, L. (1995). "Gestural syllable position effects in American English," in *Producing Speech: Contemporary Issues*, edited by F. Bell-Berti and L. J. Raphael (AIP Press, New York), pp. 19–33.
- Delattre, P., and Freeman, D. C. (1968). "A dialect study of American r's by x-ray motion picture," *Linguistics* **44**, 29–68.
- Dobson, E. J. (1957). *English Pronunciation 1500–1700*, Vol. II (Oxford University Press, London), pp. 445–1073.
- Drager, K., and Hay, J. (2012). "Exploiting random intercepts: Two case studies in sociophonetics," *Lang. Var. Change* **24**, 59–78.
- Gamer, M., Lemon, J., Fellows, I., and Singh, P. (2012). "irr: Various coefficients of interrater reliability and agreement. R package version 0.84."
- Hay, J. (2011). "Statistical analysis," in *Sociophonetics: A Student's Guide*, edited by M. DiPaolo and M. Yaeger-Dror (Routledge, London), Chap. 16, pp. 198–215.
- Hayes, A. F., and Krippendorf, K. (2007). "Answering the call for a standard reliability measure for coding data," *Commun. Methods Meas.* **1**, 77–89.
- Heselwood, B. (2009). "Rhoticity without F3: Lowpass filtering, F1-F2 relations and the perception of rhoticity in 'NORTH-FORCE,' 'START' and 'NURSE' words," in *Leeds Working Papers in Linguistics and Phonetics*, Vol. 14 (University of Leeds, Leeds), pp. 49–64.
- Jauriberry, T., Sock, R., Hamm, A., and Pukli, M. (2012). "Rhoticity et derhoticisation en anglais ecossais d'Ayrshire" ("Rhoticity and derhoticisation in the Scottish English of Ayrshire"), in *Proceedings of the Joint Conference JEP-TALN-RECITAL*, Grenoble, France (June), pp. 89–96.
- Krakow, R. A. (1989). "Physiological organization of syllables: A review," *J. Phonetics* **1**, 23–54.
- Lawson, E., Scobbie, J. M., and Stuart-Smith, J. (2011a). "The social stratification of tongue shape for postvocalic /r/ in Scottish English," *J. Sociolinguistics* **15**, 256–268.
- Lawson, E., Scobbie, J. M., and Stuart-Smith, J. (2011b). "A single case study of articulatory adaptation during acoustic mimicry," in *Proceedings of the International Congress of Phonetic Sciences XVII*, Hong Kong (August 17–21), pp. 1170–1173.
- Lawson, E., Scobbie, J. M., and Stuart-Smith, J. (2013). "Bunched /r/ promotes vowel merger to schwar: An ultrasound tongue imaging study of Scottish sociophonetic variation," *J. Phonetics* **41**, 198–210.
- Lawson, E., Scobbie, J. M., and Stuart-Smith, J. (2014a). "A socio-articulatory study of Scottish rhoticity," in *Sociolinguistics in Scotland*, edited by R. Lawson (Palgrave Macmillan, London), pp. 53–78.
- Lawson, E., Stuart-Smith, J., and Scobbie, J. M. (2014b). "A mimicry study of adaptation towards socially-salient tongue shape variants," in University of Pennsylvania Working Papers in Linguistics, Vol. 20, Article 12.
- Lennon, R. (2013). "The effect of experience in cross-dialect perception: Parsing /r/ in Glaswegian," Unpublished dissertation, submitted in partial fulfilment of the requirements for the degree of Master of Science in English Language and Linguistics in the School of Critical Studies, University of Glasgow.
- Lennon, R., Smith, R., and Stuart-Smith, J. (2015). "An acoustic investigation of postvocalic /r/ variants in two sociolects of Glaswegian," in *Proceedings of the International Congress of Phonetic Sciences*, Glasgow (August 10–15), Vol. XVIII.
- Lindau, M. (1985). "The story of /r/," in *Phonetic Linguistics: Essays in Honour of Peter Ladefoged*, edited by V. Fromkin (Academic Press, Orlando), pp. 157–168.
- Macafee, C. (1983). *Glasgow. Varieties of English Around the World, Text Series T3* (Benjamins, Amsterdam), pp. 1–167.
- Mielke, J., Baker, A., and Archangeli, D. (2010). "Variability and homogeneity in American English /l/ allophony and /s/ retraction," *Lab. Phonology* **10**, 699–730.
- Ohala, J. J. (1981). "The listener as a source of sound change," in *Papers from the Parasession on Language and Behaviour*, edited by C. S. Masek and R. A. Hendrick (Chicago Linguistic Society, Chicago, IL), pp. 178–203.
- R Core Team (2013). "R: A language and environment for statistical computing." [Computer program]. Version 3.3.2. <https://cran.r-project.org/>.
- Recasens, D., and Farnetani, E. (1994). "Spatiotemporal properties of different allophones of /l/: Phonological implications," in *Phonologica 1992: Proceedings of the 7th International Phonology Meeting*, pp. 195–204.
- Romaine, S. (1978). "Postvocalic /r/ in Scottish English: Sound change in progress," in *Sociolinguistic Patterns in British English*, edited by P. Trudgill (Edward Arnold, London), pp. 144–158.
- Scobbie, J. M., and Pouplier, M. (2010). "The role of syllable structure in external sandhi: An EPG study of vocalisation and retraction in word-final English /l/," *J. Phonetics* **2**, 240–259.
- Scobbie, J. M., Sebrege, K., and Stuart-Smith, J. (2009). "Dutch rhotic allophony, coda weakening, and the phonetics-phonology interface," in QMU Speech Science Research Centre Working Paper (QMU, Edinburgh), pp. 3–24.
- Scobbie, J. M., Turk, A., and Hewlett, N. (1999). "Morphemes, phonetics, and lexical items: The case of the Scottish Vowel Length Rule," in *Proceedings of the XIVth International Congress of Phonetic Sciences*, San Francisco, CA (August 1–7), pp. 1617–1620.
- Scobbie, J. M., Wrench, A. A., and van der Linden, M. (2008). "Head-probe stabilisation in ultrasound tongue imaging using a headset to permit natural head movement," in *Proceedings of the 8th International Seminar on Speech Production*, Strasbourg, France (December 8–12), pp. 373–376.
- Speitel, H. H., and Johnston, P. A. (1983). "A sociolinguistic investigation of Edinburgh Speech," Unpublished ESRC final report.
- Sproat, R., and Fujimura, O. (1993). "Allophonic variation in English /l/ and its implications for phonetic implementation," *J. Phonetics* **21**, 292–311.
- Stuart-Smith, J. (2003). "The phonology of modern urban Scots," in *The Edinburgh Companion to Scots*, edited by J. Corbett, J. D. McClure, and J. Stuart-Smith (Edinburgh University Press, Edinburgh, United Kingdom), pp. 110–137.
- Stuart-Smith, J. (2007a). "Empirical evidence for gendered speech production: /s/ in Glaswegian," in *Laboratory Phonology 9*, edited by J. Cole and J. Hualde (Mouton de Gruyter, Berlin), pp. 65–86.
- Stuart-Smith, J. (2007b). "A sociophonetic investigation of postvocalic /r/ in Glaswegian adolescents," in *Proceedings of the International Congress of Phonetic Sciences XVI*, Saarbrücken, Germany (August 6–10), pp. 1449–1452.
- Taylor, C. (2017). "The reliability of free school meal eligibility as a measure of socio-economic disadvantage: Evidence from the millennium cohort study in Wales," *British J. Educat. Studies* **66**, 1–23.
- Turton, D. (2014). "Some /l/s are darker than others; Accounting for variation in English /l/ with ultrasound tongue imaging," University of Pennsylvania Working Papers in Linguistics, Vol. 2, Article 21.
- Twist, A., Mielke, J., and Archangeli, D. (2007). "Are 'covert' /r/ allophones really indistinguishable?," in University of Pennsylvania Working Papers in Linguistics, Vol. 13/2, Article 16.
- Wells, J. C. (1982a). *Accents of English 3: Beyond the British Isles* (Cambridge University Press, Cambridge).
- Wells, J. C. (1982b). *Accents of English 1: Introduction* (Cambridge University Press, Cambridge).
- Wrench, A. (2012). "Articulate assistant advanced user guide," [Computer Program]. Version 2.16.12. <http://www.articulateinstruments.com/> (Last viewed July 24, 2015).
- Zhang, Z., Boyce, S., Espy-Wilson, C., and Tiede, M. (2003). "Acoustic strategies for production of American English 'retroflex' /r/," in *Proceedings of the International Congress of Phonetic Sciences XV*, Barcelona, Spain (August 3–9), Universitat Autònoma de Barcelona, pp. 1125–1128.
- Zhou, X., Espy-Wilson, C. Y., Boyce, S., Tiede, M., Holland, C., and Choe, A. (2008). "A magnetic resonance imaging-based articulatory and acoustic study of 'retroflex' and 'bunched' American English /r/," *J. Acoust. Soc. Am.* **123**, 4466–4481.