Title: Lenition and fortition of /r/ in utterance-final position, an ultrasound tongue imaging study of lingual gesture timing in spontaneous speech.

Eleanor Lawson[1], Jane Stuart-Smith[2]

[1]*CASL Research Centre, Speech and Hearing Sciences, Queen Margaret University, Edinburgh, EH21 6UU, U.K.*

[2]*GULP / English Language & Linguistics, 12 University Gardens, University of Glasgow, Glasgow, G12 8QQ, U.K.*

Abstract

The most fundamental division in English dialects is the rhotic/non-rhotic division. The mechanisms of historical /r/-loss sound change are not well understood, but studying a contemporary /r/-loss sound change in a rhotic variety of English can provide new insights. We know that /r/ weakening in contemporary Scottish English is a gesture-timing based phenomenon and that it is socially indexical, but we have no phonetic explanation for the predominance of weak /r/ variants in utterance-final position. Using a socially-stratified conversational ultrasound tongue imaging speech corpus, this study investigates the effects of boundary context, along with other linguistic and social factors such as syllable stress, following-consonant place and social class, on lingual gesture timing in /r/ and strength of rhoticity. Mixed-effects modelling identified that utterance-final context conditions greater anterior lingual gesture delay in /r/ and weaker-sounding /r/s, but only in working-class speech. Middle-class speech shows no anterior lingual gesture delay for /r/ in utterance-final position and /r/ is audibly strengthened in this position. It is unclear whether this divergence is due to variation in underlying tongue shape for /r/ in these social-class communities, or whether utterance-final position provides a key location for the performance of social class using salient variants of /r/.

Keywords: Articulatory Phonetics; rhoticity; lenition; fortition; sound change; Sociophonetics

This is a peer-reviewed, accepted author manuscript of the following article: Lawson, E., & Stuart-Smith, J. (2021). Lenition and fortition of /r/ in utterance-final position, an ultrasound tongue imaging study of lingual gesture timing in spontaneous speech. Journal of Phonetics, 86, Article 101053. Advance online publication. https://doi.org/10.1016/j.wocn.2021.101053

Lenition and fortition of /r/ in utterance-final position, an ultrasound tongue imaging study of lingual gesture timing in spontaneous speech.

1.0 Introduction

World Englishes can be split into two categories: rhotic and nonrhotic (Wells, 1982a, §1.3.3). In rhotic varieties, postvocalic /r/ is pronounced in all contexts. In non-rhotic varieties, postvocalic /r/ is pronounced only in prevocalic contexts, e.g. in "far away", but not in "far flung". Historical loss of /r/ in Anglo English in the eighteenth century significantly expanded the vowel systems of varieties in which it has occurred, phonemicising prerhotic vowels that had undergone phonetic changes due to the coarticulatory effects of the following /r/ (Dobson, 1957; Wells, 1982a). This phenomenon makes rhoticity/nonrhoticity perhaps the most fundamental systemic phonological division in English (Dobson, 1957; Wells, 1982a; McMahon, Heggarty, McMahon and Maguire, 2007; Maguire, McMahon, Heggarty and Dediu, 2010).

It is difficult to determine how loss of postvocalic /r/ began in historical Englishes due to a lack of contemporary evidence. Historical accounts of loss of postvocalic /r/ in the eighteenth and nineteenth centuries are rare, intriguing, and difficult to decipher. The eighteenth century elocutionist John Walker, describes /r/ in England, and particularly in London, as being pronounced "in the throat" and involving a "vibration of the lower part of the tongue near the root" (Walker, 1791, p.50), in contrast to the Irish trilled /r/ "jarring of the tip of the tongue against the roof of the mouth near the fore teeth" (ibid.). There is also a late nineteenth century American account of loss of postvocalic /r/ in Virginia where "The vocal organs assume the proper position for pronouncing the soft (i.e. approximant) r and then stop before producing the sound." (Primer, 1890, p.199).

Today, in Central Scottish English, we see derhoticisation occurring again; weakening of postvocalic /r/ in working-class Scottish English speech has been the subject of sociolinguistic research for over half a century, with real-time evidence showing weakened /r/ for at least a century (see §1.1). It has been established that working-class Scottish English derhoticisation is a vernacular sound change, and a change from below (Labov 2001), not brought about through contact with speakers of non-rhotic varieties of British Isles English (Speitel and Johnston, 1983). Articulatory analysis of word-list data, using ultrasound tongue imaging (UTI), shows that lingual gesture delay is a key feature of /r/ weakening, leading to weakly rhotic variants, and even apparent deletion, at the auditory level (Lawson, Stuart-Smith and Scobbie, 2018). UTI has proved to be an excellent method for the study of socially-indexical derhoticisation, as it is physically non-invasive, does not appear to affect speech style any more than audio recording equipment, and permits recording of greater numbers of informants than other articulatory techniques (Lawson, Stuart-Smith and Scobbie, 2008). However, to date, UTI-based studies have focused on word and phrase-list data, while spontaneous-speech data has been difficult to obtain, due to large file sizes involved and the need for regular file saving in order to avoid loss of data during recording. A focus on word-list speech style means that the impact of prosody on /r/ weakening has not been studied. Evidence from auditory studies of working-class Scottish English spontaneous speech, suggest that the position of the /r/ segment within the utterance plays a significant role in lenition, with utterance-final /r/s showing the greatest degree of audible weakening (see §1.1). The current study represents the first attempt to study the effects of utterance position on rhotic gesture timing, by making use of an audio-UTI conversational-speech dataset recorded for an accommodation study. This paper outlines the conversational speech data-collection method and assesses the impact that social and linguistic factors have on both lingual gesture timing in /r/ and audible rhotic strength. In doing so, it provides a more comprehensive understanding of the processes involved in this important sound change, and a better understanding of how prosody, speech articulation and social factors interact in sound change processes.

1.1 Derhoticisation in Scottish English

Scottish English is usually thought of as “firmly rhotic” (Wells, 1982b, p.140) with rhoticity being an important marker of Scottish identity. The English Dialects App (EDA) survey shows a sharp discontinuity around the Scottish-English border, in an analysis of responses to the question, “I pronounce the word ‘arm’ as: (1) [r] (2) no [r]” (Leeman, Kolly and Britain, 2018, Figure 16). In the past, /r/-lessness in Scottish English was associated with aspirational anglicisation of middle-class Scottish English speech, often acquired through elocution lessons. From the middle of the twentieth century, this kind of Scottish middle-class nonrhoticity became unfashionable and was in decline (Johnston, 1985). However, a large body of, mainly sociolinguistic, research into postvocalic /r/ in the most populous region of Scotland, the lowland Central Belt, see recording locations in Figure 1, has shown changes in the rhoticity of vernacular speech across that area, with working-class (WC) speakers often producing weakly-rhotic utterances:

Edinburgh – (Romaine, 1978; Speitel and Johnston, 1983)
Glasgow – (Macafee, 1983; Stuart-Smith, 2007)
Livingston – (Lawson et al, 2008)
Ayr – (Jauriberry, Sock, Hamm and Pukli, 2012)

Figure 1: Map of the Scotland, showing locations across the Central Belt where derhoticisation has been studied.

Weakening of postvocalic /r/ can result in tokens that sound completely /r/-less, but more often result in tokens that are difficult to classify as /r/-ful, or /r/-less; these are often labelled “derhoticised”. Reminiscent of Walker’s description of eighteenth century Anglo-English postvocalic /r/ being pronounced “in the throat” (Walker, 1791, p.50), present-day Central Scottish English words containing derhoticised /r/s often have pharyngealised nuclear vowels, while having no unambiguously audible rhotic segment, e.g. [fɑ$^{(ʕ)}$][1] “far”, [biɐ$^{(ʕ)}$] “beer”, [bʌ$^{(ʕ)}$d] “bird” (Speitel and Johnston, 1983; Stuart-Smith, 2003).

Studies using auditory methods to classify /r/ variants all point towards the gradience of the sound change, and the subtlety of remaining acoustic cues, e.g. a perceptual study of derhoticised postvocalic /r/ variants in Glaswegian speech showed that while nonrhotic speakers from Cambridge (southern England) who are unfamiliar with Glaswegian English, performed at below chance level when discriminating between Glaswegian /r/-ful and /r/-less minimal pairs, Glaswegian speakers easily discriminated between them (Lennon, 2013), picking up on subtle acoustic cues that are socially salient in their speech community. However, even where listeners are local and phonetically trained, there is often a lack of

[1] N.B. It is difficult to represent derhoticised variants with conventional segmental phonetic transcription methods.

agreement as to whether /r/ is present or absent (Stuart-Smith, 2007; Lawson, Scobbie and Stuart-Smith, 2014a).

A further feature of Scottish English derhoticisation that has been noted again and again is the importance of utterance position. Studies of the postvocalic-/r/ weakening phenomenon in working-class Central Belt speech, using recordings going back as far as 1916, identify that utterance-final position favours /r/ weakening, and apparent deletion. In a study of /r/ in the speech of working-class Edinburgh children, Romaine identified that apparent deletion of /r/ was most common in utterance-final position, particularly for male speakers (Romaine, 1978). In their socially-stratified study of Edinburgh adult speech, Speitel and Johnston (1983) noted that partial devoicing of postvocalic /r/ was common in working-class speech in utterance-final position, and also that derhoticised variants were more common in this position. In an analysis of postvocalic /r/ data collected in the late 90s in Glasgow, Stuart-Smith (2003) identified that apparent deletion of postvocalic /r/ was most common in unstressed word- and utterance-final position. Lawson et al (2008) found apparent deletion of /r/ most often in utterance-final position, particularly in unstressed syllables and following high vowels. In his study of western Central Belt speech in Ayr, Jauriberry also found the greatest percentage of deleted /r/ in utterance-final position (Jauriberry et al, 2012). Finally, a study of early 20th century postvocalic /r/, based on recordings made on shellac discs of mainly working-class Central Belt Scottish males in their 20s (recorded by German dialectologist Willhelm Doegen in prisoner of war camps) showed that utterance-final position conditioned the greatest percentage of weak (derhoticised and apparently deleted) postvocalic /r/ variants (Stuart-Smith and Lawson, 2017).

1.2 Articulatory correlates of postvocalic/r/ weakening: gesture timing

Understanding Scottish English postvocalic /r/ weakening and loss has proved difficult without recourse to articulatory data (Stuart-Smith, 2007). The two most likely articulatory causes for gradient /r/ weakening and loss in this community are variation in (1) gesture magnitude and (2) gesture timing. A reduction in gesture magnitude can lead to more vocalised /r/ variants and evidence suggests that lingual consonants in syllable-coda position tend to have a reduced magnitude compared to syllable-onset position, see (Browman and Goldstein, 1995; Giles and Moll, 1975; Kochetov, 2006; Lawson, Leplatre, Stuart-Smith and Scobbie, 2019). However, an increasing body of evidence points to the importance of variation in gesture timing in weakening of postvocalic liquid consonants. Gesture timing has been measured using cineradiography (Sproat and Fujimura, 1993); electropalatography (EPG) (Recasens, and Farnetani, 1994); electromagmentic articulometry (EMA) (Kochetov, 2006) and ultrasound tongue imaging (UTI) (Gick, Campbell, Oh and Tamburri-Watt,2006; Lawson et al, 2018; Lawson and Stuart-Smith, 2019). In the current study, we use UTI to study gesture timing.

There are at least two ways in which gesture timing in liquid consonants (including rhotics) can be measured, and they depend on the phonetic contexts of the liquid and provide different types of information. Where liquids are flanked by high vowels, e.g. "nea*r* east", anterior and posterior lingual gesture maxima can be identified and annotated (e.g. tongue-tip raising maximum and tongue-root retraction maximum), then the intergestural duration (the time between the two gesture maxima) can be measured, see (Sproat and Fujimura, 1993; Lawson and Stuart-Smith, 2019). Increased separation of anterior and posterior gesture maxima in another liquid consonant, /l/, has been shown in to correlate with acoustic darkness and vocalisation (Sproat and Fujimura, 1993).

Where /r/ occurs in a variety of vowel contexts, as happens in unstructured, spontaneous-speech datasets, it is less easy to identify and annotate the posterior root-retraction gestures. In these cases, only the anterior gesture maximum can be reliably identified. Nevertheless, identification of only the maximum of the anterior gesture can still provide useful information on /r/ weakening, as the temporal position of the anterior gesture can be studied relative to other temporal speech events that could potentially cause audible masking of the rhotic gesture, e.g. voicing offset, or the onset of a following consonant, see (Lawson et al, 2018). Lawson et al (2018) used this anterior-lingual timing measurement method to study rhotic gesture timing and identified that the more delayed the anterior tongue gesture was, relative to voicing offset or final (nonlingual) consonant onset, the higher F3, the lower F2 and the less /r/-ful sounding the token (Lawson et al, 2018). Working-class speakers in that study, whose /r/s were rated significantly less rhotic-sounding than those of middle-class speakers, were found to produce the maxima of their anterior /r/ gestures, on average, after the offset of voicing, or after the onset of a following labial consonant. Middle-class speakers in the same study produced tokens that were audibly more rhotic-sounding and their anterior tongue-gesture maxima almost always occurred before voicing offset/following-consonant onset. These findings help explain the presence of tokens that are ambiguously rhotic in Central Belt Scottish English speech; the /r/ is still present at the articulatory level, but is audibly masked to varying extents.

Measures of intergestural separation (Sproat and Fujimura, 1993; Gick et al, 2006), or anterior gesture timing relative to other temporal speech events (Lawson et al, 2018), measure the same underlying phenomenon in different ways and provide different insights into the lenition process. However, one issue with Lawson et al (2018) is that it considered word-list speech only, where /r/ was effectively always in utterance-final position, and post-rhotic consonants in the study were always non-lingual. Therefore, there was no possibility to investigate the phenomenon of increased derhoticisation in utterance-final position, as no tokens of /r/ in non-utterance-final position were available for comparison.

We also could not study the effects of a variety of post-rhotic consonants on the lingual gesture timing of /r/.

1.2.1 Social stratification of /r/ tongue shape in Scottish English

In addition to variation in gesture timing, UTI-based research has uncovered another fundamental type of socially-stratified articulatory variation for Scottish English /r/, namely tongue-shape variation. Using a socially-stratified ultrasound corpus of adolescent Central Belt speech, Lawson, Scobbie and Stuart-Smith (2011) showed that working-class speakers in the corpus tend to produce coda /r/ with a raised tongue tip, while middle-class speakers produced bunched postvocalic /r/ variants with a lowered tongue tip and raised tongue front or dorsum. This finding was subsequently replicated in (Lawson et al, 2014a; Lawson, Scobbie and Stuart-Smith, 2014b). While bunched/retroflex variation in /r/ has also been identified in American English, this variation has been shown to be idiosyncratic (Delattre and Freeman, 1968; Mielke, Baker and Archangeli, 2010), rather than socially conditioned. The origins of bunched /r/ in middle-class Scottish English are unclear; however, one potential explanation is that bunched variants produce a stronger auditory percept of rhoticity, e.g. Delattre and Freeman (1968) remarked that their American English /r/ type 4 ‘dorsal bunched with dip’ produced the “strongest auditory impression” of /r/ (Delattre and Freeman 1968, p. 64). Use of bunched /r/ could therefore be a means of strengthening /r/, as a deliberate contrast to working-class derhoticisation. Additionally, it has been identified that use of postvocalic bunched /r/ causes a three-way *fir*, *fur*, *fern* vowel merger through coarticulation (Lawson, Scobbie and Stuart-Smith, 2013), the so-called NURSE merger (Wells, 1982a), which is not a traditional feature of Scottish English and is found in middle-class, but not in vernacular Scottish English. There is evidence to suggest that acquisition of this merger, and /r/ variants that promoted it, were encouraged through elocution in early-to-mid twentieth century Scotland (McAllister, 1938). It is likely that postvocalic /r/ tongue shape affects tongue-gesture timing, as variants that involve formation of a constriction using the tongue body are more likely to have maxima close to the syllable nucleus, whereas /r/ variants that involve formation of a constriction with the tongue tip allow more scope for anterior gesture drift towards, and past, syllable margins.

1.3 The effect of linguistic boundaries on liquid consonant gesture timing.

The notion that syllabic duration is manipulated by speakers to reflect syntactic structure is well established. In her 1973 paper, Lehiste showed that metrical feet were around one third longer in duration in utterance-final position, than in other positions. She also showed pre-boundary foot lengthening being used to disambiguate meaning in lexically-identical sentences (Lehiste, 1973). Wightman, Shattuck-Hufnagel, Ostendorf and Price (1992) showed that pre-boundary lengthening is confined to the rime of the syllable preceding each prosodic boundary, with the duration of all rime constituents increasing in relation to the strength of the following boundary (Wightman et al, 1992). Pre-boundary lengthening has been found to affect the duration of consonant production, slowing it down, as shown in Fougeron and Keating's EPG study of [n] in reiterant utterances (repetition of the syllable "no" in place of real speech syllables) and in Krivokapic and Byrd's EMA study of consonant-to-consonant (and vowel-to-vowel) durational measures across syntactic boundaries of different strengths (Krivokapić and Byrd, 2012). Sproat and Fujimura's (1993) cineradiographic study of American English /l/ found a correlation between linguistic boundary strength, preboundary syllable-length, intergestural duration (temporal duration between tongue-tip and dorsum maxima) and the darkness of /l/ (measured acoustically as distance between F1 and F2), with boundary strength predicting intergestural duration and the greatest inter-gestural separation occurring before a major intonation boundary. An EPG-based study of gesture timing in /l/ in Catalan and English phrase-list recordings showed that, in utterance-final position, anterior lingual gestures for consonants are not only slowed down, they can also be delayed to the point that they are inaudible, occurring after voicing has ended (Recasens and Farnetani, 1994). If anterior rhotic gesture delay is augmented by prosodic factors in Scottish English, then word-list recordings will emphasise postvocalic /r/ weakening, as all tokens of postvocalic /r/ in word lists effectively occur before a major prosodic boundary, as pointed out by (Lehiste, 1973). The auditory studies of /r/ in spontaneous speech, detailed in §1.1, showing utterance-final position as a key location for /r/-weakening and loss in working-class Scottish English, suggest that anterior rhotic gesture delay may be increased by prosody in this variety.

The aim of the present study, therefore, is to gain a better understanding of the impact of the utterance-final position on the gesture timing of rhotic segments and their auditory quality, taking into account other potential factors affecting Scottish English /r/ weakening; social factors such as class and sex, and linguistic factors such as stress, presence of a post-rhotic consonant and place of articulation of the post-rhotic consonant, using a spontaneous speech audio-articulatory dataset. In doing so we may also gain a better understanding of the historical /r/-loss process, and a better understanding of how articulatory, prosodic and social factors interact to produce sound change. The study's research questions are as follows:

(1) What impact does utterance-final position have on the timing and auditory quality of postvocalic /r/ in Scottish English?
(2) How does utterance position interact with other linguistic factors and with social factors in the /r/ weakening process?

2. Method

2.1 Informants

Adolescents were recorded for this study, partly due to the fact that better-quality ultrasound images are obtained from younger speakers. Using school-age participants also helped with the social stratification of the dataset, as pupils could be recruited from schools with affluent or deprived catchment areas. Adolescents, aged 12-13, were recruited from two schools (located approximately 7km apart); one in a deprived catchment area in Glasgow, and one in an affluent catchment area in Glasgow in 2015[2]. The study aimed to recruit 4 males and 4 females from each school; however, one male and one female student from the school with the deprived catchment area did not attend on the day of recording. Two participants volunteered to be recorded twice in conversation with their peers, but only the data from their first recording session was analysed. Participants were recruited by teachers in their schools and were given gift vouchers to thank them for participating in the study.

2.2. Equipment, set up and materials

Informants were recorded in inter-social-class pairs, carrying out a spot-the-difference conversational task with audio and ultrasound tongue imaging (UTI) in an IAC sound-attenuated recording booth at the University of Glasgow. The equipment set-up is represented in Figure 2 below. The researcher, and all noisemaking equipment such as the ultrasound machines and laptops were located outside of the recording booth. Participants were fitted with stabilising headsets, holding the ultrasound probe in place underneath the chin, to eliminate coronal and axial rotation, and minimise sagittal rotation movement of the ultrasound probe in relation to the head (Scobbie, Wrench and Van der Linden, 2008). Audio recordings were sampled at 22kHz and were made using Beyer-Dynamic Opus 55 neck-worn microphones. Two Mindray DP2200 ultrasound machines, set to NTSC video format, created UTI video at a target rate of 29.97fps. The frame rate of the UTI video was doubled to 59.94fps by deinterlacing each video frame post hoc. UTI recordings were made using *Articulate Assistant Advanced* software (AAA) (Wrench, 2012) on two separate Toshiba Tecra laptops with dedicated video cards. Audio-

[2]School 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.

ultrasound resynchronisation and frame-rate establishment for these recordings was achieved using an electronic clapperboard system (SynchBrightUp), which adds a bright square to the video signal and a tone and pulses to the audio signal the instant they are created, and these two signals are synchronised automatically post hoc. The resynchronisation process is set out in (Lawson et al, 2018): §II.D).

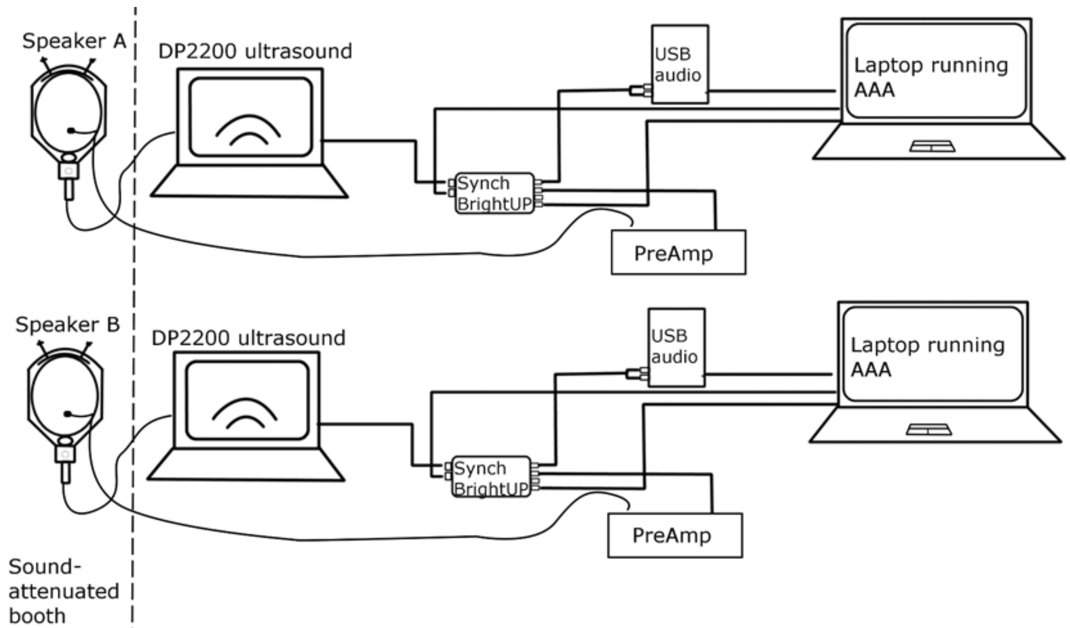


Figure 2: Experimental equipment set up for recording two speakers in conversation with synchronized audio and ultrasound tongue imaging. Broken line indicates the recording-booth boundary.

At the time of the recordings, it was not possible to continuously record speech using the AAA ultrasound-recording system, therefore the conversational recordings that took place were automatically sampled, with 10s recording time, followed by 10s of saving time, and so on. The mean duration of recorded speech was therefore 11m 30s (+/- s.d. 4m 23s), though participants conversed for double this time. Speakers produced 8 - 43 words containing postvocalic /r/, mean 24, see Table 1, for lexical-item frequencies.

One of the original aims of data collection was to study accommodation in spontaneous speech, therefore each conversational dyad consisted of one working-class and one middle-class adolescent. With the exception of one working-class male speaker, who produced only 8 /r/ tokens, participants showed good levels of engagement with one another, despite never having met their interlocutors before.

Participants were each given a picture containing items designed to elicit examples of postvocalic /r/ in stressed and unstressed position, see Figure 3. They were told that there were 26 differences to spot between each picture, that they should take turns to ask questions about their interlocutor's picture, circle the areas on the image where there were differences, and that they should not show their picture to their conversational partner. They were told that the researcher would be monitoring their speech wave forms, but could not hear their conversations while they were speaking. Typically, after a few minutes, participants would report that the task was too difficult and were prompted to look more carefully, after which they would attempt to find most of the differences.

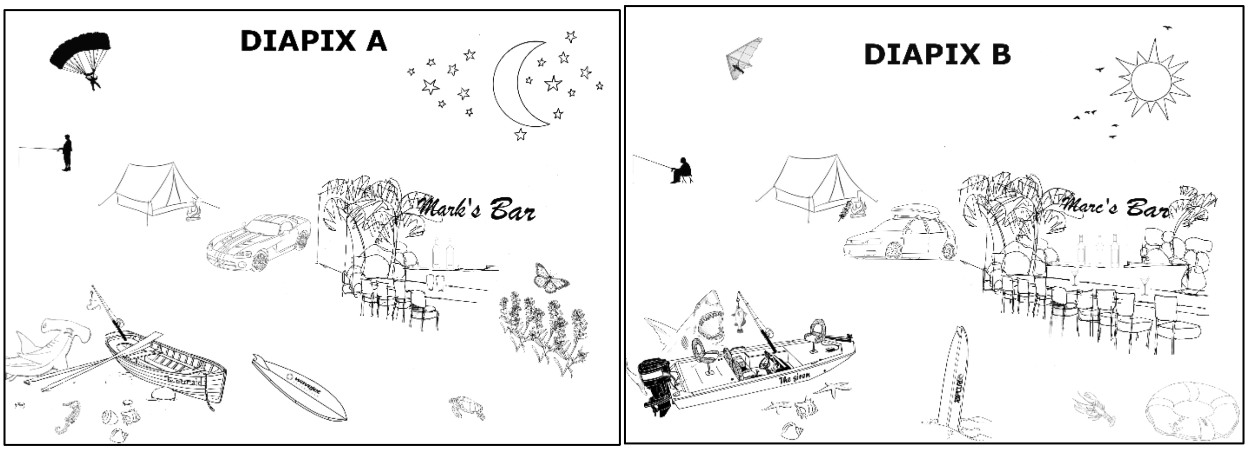


Figure 3: Spot-the-difference pictures A and B.

Table 1: /r/ word frequencies, N=351

Lexical items	No. of occurrences
bar	23
your	21
four	19
car	17
Mark's/Marc’s; more; there’s	11
chair(s); shark(s); surf; star(s)	10
board; circled; there; yours	9
or, circle(s), seahorse	8
corner; fourteen; marshmallow(s)	7
turtle; birds	6
thirteen; where	5
fire; hammerhead; hard; lobster; near; starfish; there’s	4
air; are; for; mirror; oars; (para)glider; worm	3
butterfly; door(s); fisherman; flowers; mermaid; normal; over; sure; they’re; weird	2
another; archy; arm; chair; counter; creatures; curl; ear; first; hammerfish; hardly; hour; Martini; matters; person; score; sore; sorta; sports; tortoise; underneath; we’re; were; works; you’re	1

2.3 Temporal annotation

Wave files were exported from AAA and syllable *rime duration* was annotated for each /r/-word token using Praat (Boersma, Weenink 2013). The syllable rime included the nuclear vowel (including any period of aspiration after onset plosive release), postvocalic /r/ and any other coda consonants following the /r/, excluding those that had been audibly resyllabified (12 tokens).

As this study involves spontaneous speech, it was not possible to select only tokens preceded by high vowels, therefore, the present study quantifies the durational difference between *rmax*, the maximum of the anterior lingual gesture in postvocalic /r/ (tip/front raising or dorsum raising, depending on the type of /r/ involved), which was visually determined using ultrasound analysis software AAA (Wrench, 2012), and either the offset of vocal-fold vibration, *voice-offset*, or following consonant onset, *C-onset*; annotated using Praat (Boersma and Weenink, 2013), both of these latter annotations identify key temporal events that could render the /r/ gesture partly, or completely, inaudible. This method of quantifying gesture *lag* was also used in (Lawson et al, 2018).

rmax was annotated in AAA by comparing tongue movement frame by frame. The *rmax* annotation was added at the first UTI video frame where the tongue tip, blade, front or dorsum (depending on whether tip-up, or bunched /r/ was produced) reached its highest position (see Figure 4.) All AAA annotations were then exported in Praat TextGrid format, along with their associated audio files in wave format, for further annotation using Praat. *V-onset*, and *voice-offset*, or *C-onset* values, and also syllable-rime onset and offset, were annotated by the first author using a combination of waveform, spectrogram and Praat tools such as the pitch and intensity trackers as a guide. There were 324 annotated /r/ tokens in the dataset, a further 27 tokens were not annotated, 10 of which were instances where there appeared to be no /r/ gesture (5 of these were produced by one speaker, GWM5 and 3 by GWM7), and 17 of which had an /r/ gesture where the maximum could not be easily identified. In the current study, intervocalic context is explicitly excluded, as it conditions a tap [ɾ] for many Scottish English speakers.

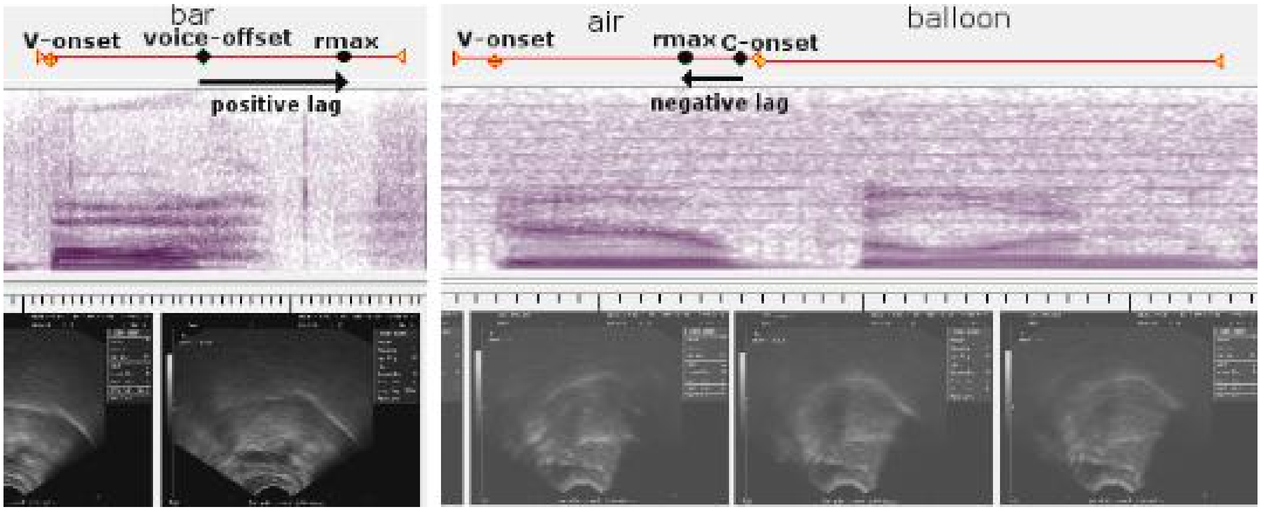


Figure 4. (Left) WC speaker GWF8, utterance-final "bar" [ba̠ʰ] showing *V-onset* (yellow diamond), *voice-offset* (black diamond) and *rmax* (black circle) annotations above the spectrogram. (Right) MC speaker (GMF9) "air balloon" [eɚbəlʉn] showing *V-onset* (yellow diamond), *rmax* (black circle) and *C-onset* (black diamond) annotations above the spectrogram. Tick marks below the spectrogram show the alignment of the UTI frames with the acoustic signal – UTI frames are 18ms apart. Below the spectrogram are sample UTI image frames showing the progression of the tongue movements through the recording.

For each postvocalic /r/ token, three of the four main temporal events listed below 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 the syllable rime.
(3) *voice-offset* - the acoustically-determined location of the offset of voicing where /r/ is not immediately followed by a consonant.
(4) *C-onset* - the acoustically-determined location of the onset of an immediately-following consonant, either in the same syllable, a following syllable, or a following word.

A raw *lag* measure was calculated by subtracting *rmax* from either voice-offset, or C-onset. This measure could be a positive value, as in Figure 4 (left), or a negative value, as in Figure 4 (right). We would predict that positive *lag* values would be associated with weaker sounding /r/s, as some of the rhotic articulation occurs after a masking event, such as voice-offset, or following-consonant onset, and /r/ would therefore be at least partially audibly masked. 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. Raw and normalised lag were highly correlated r_P=0.81, p<0.001.

2.4 Linguistic boundary annotation

There are different approaches to linguistic boundary annotation. Many studies have focused on the boundaries between prosodic units, rather than syntactic or morphological units. Both (Price, Ostendorf Shattuck-Hufnagel and Fong, 1991) and (Wightman, Shattuck-Hufnagel, Ostendorf and Price, 1992) use a 7-level break index to study read-speech corpora, with annotators recording prosodic boundary strengths between words, from the prosodic-word level, to the sentence-boundary level. The nature of the sampled spontaneous speech dataset used in the current paper, where utterances are often truncated, makes this kind of analysis difficult. Sproat and Fujimura (1993) used an approach encompassing morphological, syntactic and prosodic boundaries, marking boundaries between morphological units, word and phrase-level boundaries, and major intonation boundaries. They organized this set of boundaries in an a priori ascending order of strength to use as factorial levels in the analysis of articulatory and acoustic measures. Again, Sproat and Fujimura (1993)'s dataset consisted of read-speech with a highly-structured set of stimuli. In the current study, a simplified set of boundary contexts was annotated, see Table 2 below. Initially, more boundary levels were coded. The *utterance-boundary* level combines both (i) pre-pausal and (ii) turn-final contexts, with the former category characterising a pause or hesitation that is not turn final with a mean 0.92s duration before the participant spoke again, and the latter category occurring when the participant completed their utterance and yielded the floor. ANOVAs determined that, for the three dependent variables analysed, there were no significant differences between pre-pausal and turn-final contexts: *rime duration* (F=3.064, p=0.0831); *normalised lag* (F=3.114, p=0.0806) and *r-index score* (F=1.532, p=0.219), so these two categories were combined into a single boundary category called "utterance boundary". Likewise, there were no significant differences for the three dependent variables analysed, between word-final contexts that were (i) phrase-medial or (ii) phrase-final: *rime duration* (F=0.898, p=0.345); *normalised lag* (F=0.422, p=0.517) and *r-index score* (F=0.765, p=0.383), therefore, these two categories were combined into a single boundary category called "word boundary".

Table 2: Linguistic contexts for /r/ tokens in the study

Boundary context	Code	example	Ns (total=311)
Utterance	Vr(C)###	"bar…" "mirror…" "starts…"	104

Word	Vr(C)##	"car doors" "cars have" "for the"	137
Syllable	Vr(C)#	"turtle" "starfish" "thirteen"	70

Temporal annotations were extracted from the TextGrid tiers using a Praat script and saved as a .csv file where social and other linguistic factors were then coded for statistical analysis with R (v 3.1.2) (Core Team 2018). Syllable stress and following-consonant place of articulation were also recorded for each postvocalic /r/ token. In Lawson et al (2018), the post-rhotic consonant was always nonlingual, but in the current study a variety of post-rhotic consonants were present. To date, we do not know how following consonants with lingual articulations affect gesture delay in /r/; however, we hypothesized that greater degrees of gesture lag are to be expected when /r/ is followed by a non-lingual consonant, or a pause. The consonants that followed /r/ tokens in the spontaneous speech dataset had a variety of places of articulation, and were collapsed into the following four place categories, relating the articulatory specification of the tongue: *coronal*: dental [ð]; alveolar [s], [t], [d], [n], [l]; postalveolar [ʃ], [ʧ], [ʤ]; *velar*: velar [k], [g], [w]; *nonlingual*: bilabial [b], [p], [m]; labiodental [f]; and glottal [h], [ʔ]. Where /r/ was not followed by a consonant, e.g. CVr words in utterance-final position, consonantal place was coded as "null". We hypothesized that where /r/ was followed by (1) no consonant, (2) a nonlingual consonant, or (3) a velar consonant, greater delays in the anterior /r/ gesture would be permitted than when /r/ was followed by a *coronal* consonant.

2.5 Audio rating analysis

Audio tokens of /r/ words were extracted from the spontaneous speech data, including tokens where no /r/ articulation had been identified. Tokens had a contextual margin of 50ms on either side of the /r/ segment. Each instance of /r/ was rated for rhoticity strength by the first author (a native speaker of Central Belt Scottish English) on a 5-point r-index using a Praat multiple forced choice (MFC) experiment. The /r/-index scale is based on a scale established in Lawson et al (2011). The /r/-index scale contains two International Phonetic Association (IPA) labels, "alveolar" and "retroflex", and is augmented with categories "no audible /r/", "derhoticised /r/" and "rhoticised vowel". The /r/-index is a purely auditory rating index, so identifying a token as "retroflex", for example, means only that the token sounds retroflex. Ultrasound was not used to confirm the presence of retroflection.

The categories here are illustrated using the word "corner" to show both stressed and unstressed contexts: (1) no audible /r/ [kʰɒnʌ] (2) derhoticised /r/ [kʰɒˤnʌˤ], (3) alveolar /r/ [kʰɒɹnʌɹ], (4) retroflex /r/ [kʰɒɻnʌɻ], (5) rhoticised vowel [kʰɒ˞nɚ]. In this /r/-index, we consider that an /r/ identified as "alveolar" sounds less strongly rhotic than an /r/ identified as "retroflex". Although rhoticisation of a vowel could be considered to be a type of consonant vocalization, resulting in a reduction in /r/ strength, vowel rhoticisation in this speech community effectively increases the duration of rhotic acoustic cues in the syllable, which strengthens the percept of rhoticity, therefore rhoticised vowel is considered to be the strongest-sounding variant in this rhotic continuum.

Tokens were presented randomly and anonymously, together with an orthographic representation of the word. An indefinite number of replays were permitted and a "can't classify" button was available. 311 of the articulatorily-annotated tokens were auditorily classifiable. The rest were not classified due to various factors such ambient noise, interlocutor speech noise, overlapping noise from the synch pulse, which was recorded on the audio channel, or just difficulty categorising; this last reason was common for short, unstressed syllables.

2.6. Statistical analysis

Linear mixed effects modelling was carried out using R (v 3.1.2) (Core Team 2018) followed by the step() function in the LmerTest package (Kuznetsova, Brockhoff and Christensen, 2017) to find the models that best fit the data. Full models are provided in footnotes, and best-fit models are given at the top of each dependent variable subsection in §3 Results. The dependent variables studied were *rime duration*, *normalised lag* and audio *r-index score*. Fixed factors used in statistical modeling were both social and linguistic and are shown with their levels and numbers of tokens in Table 3 below. Random intercepts (1) SPEAKER and (2) WORD were also included in all full models.

Table 3: List of fixed factors used in statistical analysis with levels and numbers of tokens per level.

Fixed factors	**Level 1**	**Level 2**	**Level 3**	**Level 4**
(1) SEX	**Male**	**Female**		
	N=137	N=174		

(2) social CLASS	**Working class**	**Middle class**		
	N=112	N=199		
(3) syllable STRESS	**stressed**	**unstressed**		
	N=246	N=65		
(4) TAUTOSYLLABIC CONSONANT	**yes** (1 or more tautosyllabic consonants following /r/)	**no** (/r/ is the final consonant in the syllable)		
	N=103	N=208		
(5) BOUNDARY context	**utterance boundary**	**word boundary**	**syllable boundary**	
	N = 70	N = 137	N = 104	
(6) following-consonant PLACE	**coronal**	**null**	**nonlingual**	**velar**
	N = 134	N = 81	N = 58	N = 38

For *rime duration*, the full model included fixed factors SEX, CLASS, STRESS and BOUNDARY, with the random intercepts, and we tested for an interaction between CLASS*BOUNDARY.

For *normalised lag* and *r-index*, initial full models tested included fixed factors: SEX, CLASS, STRESS, BOUNDARY and TAUTOSYLLABIC CONSONANT. We also tested for an interaction between CLASS and BOUNDARY. An alternative pair of models were also tested for *normalised lag* and *r-index*, substituting BOUNDARY context and TAUTOSYLLABIC CONSONANT, with the factor following-consonant PLACE, testing for an interaction between CLASS and PLACE. These initial and alternative models were considered, because we were interested in the effects of both boundary CONTEXT and following consonant PLACE, but wished to avoid collinearity - the PLACE level "null" overlapped to a great extent with the CONTEXT level "utterance boundary". These alternative models were compared for parsimony of fit, using the AIC function of the R stats package (Core Team, 2018). The emmeans package (Lenth, 2019) was used to carry out Tukey posthoc tests on significant factors for each model.

A Spearman's rank correlational analysis was also carried out between the articulatory dependent variable, *normalised lag*, and the auditory variable, *r-index score*.

3. Results

3.2 Mixed effects modelling

3.2.1 rime duration

For the *rime duration* dependent variable, the final[3] model was:

rime_duration ~ class + stress + boundary + class:boundary + (1 | speaker) + (1 | rword)

The fixed factor STRESS was significant, $F(1,238) = 32.768$, $p<0.001$, unstressed rimes (estimated mean 147ms) were 81ms shorter than stressed rimes (estimated mean 228ms). There was also an interaction between BOUNDARY and CLASS, $F(2, 287.502) = 4.848$, $p=0.0085$. Figure 5 shows that a significant difference was found between working-class and middle-class speakers' rime durations in utterance-boundary context only $t(34.9)=3.962$, $p=0.0044$. While middle-class speakers have adjusted mean utterance-final rimes 0.303s long (s.e. +/-15ms), working-class speakers' adjusted mean utterance-final rimes were 71ms shorter, at 0.232s (s.e. +/-17ms). Other significant variation between BOUNDARY context by social CLASS is listed in Table 4, which shows that there were significant durational difference between syllable rimes across the three boundary contexts for both social classes, with rime duration increasing significantly by BOUNDARY context as follows: syllable < word < utterance.

[3] Full model before step() comparison: rime_duration ~ sex + class + stress + boundary + class*boundary + (1|speaker)+ (1|rword)

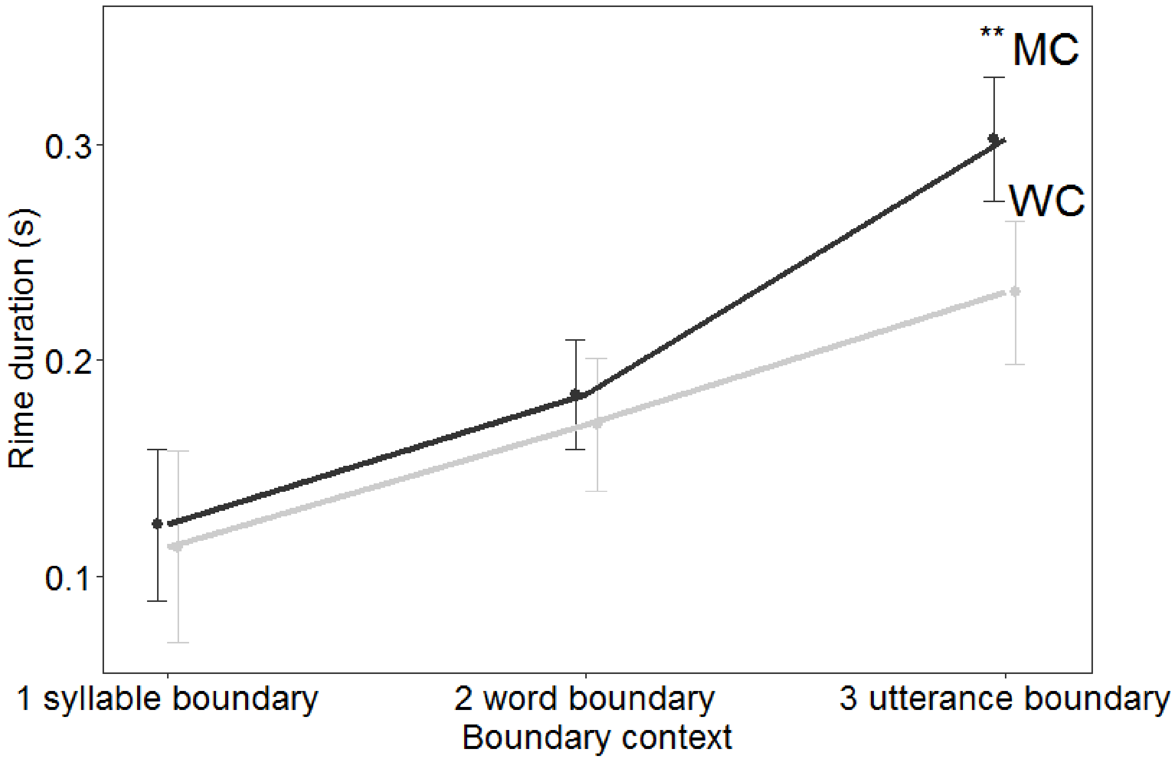


Figure 5: Plot showing estimated means for *rime duration* by BOUNDARY context, N=311.

Table 4: Other significant variation between BOUNDARY context by social CLASS.

MC syllable boundary – utterance boundary ***	$t(165.8) = 8.841$ $p < 0.0001$
MC word boundary – utterance boundary ***	$t(293.4) = 8.849$ $p < 0.0001$
WC syllable boundary – utterance boundary ***	$t(232.5) = 4.605$ $p = 0.0001$
WC word boundary – utterance boundary **	$t(286.4) = 3.650$ $p = 0.0042$

These findings, based on spontaneous-speech data, confirm the boundary-strength-dependent lengthening of syllable-rimes found by (Sproat and Fujimura 1993; Wightman et al, 1992) and reaffirm that utterance-final position conditions significant syllable-rime lengthening (see Table 4). However, the social factor CLASS also has an impact on utterance-final rime lengthening, with utterance-final rimes being significantly shorter when produced by working-class speakers, than when produced by middle-class speakers (see Figure 5).

3.2.2 Normalised lag

The *normalised lag* measure quantifies temporal distance between the maximum of the anterior lingual rhotic gesture (tongue tip, front, or dorsum, depending on the kind of /r/ produced) and either voicing offset, or the onset of a following consonant. Negative *normalised lag* values indicate that the maximum of the anterior rhotic gesture occurred before voicing offset, or onset of the following consonant, while positive values indicate that the anterior rhotic gesture occurred after voicing offset, or during the articulation of a following consonant.

Two separate models were tested for the *normalised lag* dependent variable, substituting BOUNDARY context and TAUTOSYLLABIC_CONSONANT in model 1, with following-consonant PLACE in model 2, to avoid collinearity, as the *utterance_final* level of the BOUNDARY factor overlaps to a great extent with the *null* level of PLACE.

3.2.2.1 Normalised lag – model 1

The first model tested for the effects of SEX, social CLASS, STRESS, BOUNDARY context, and TAUTOSYLLABIC_CONSONANT; testing for interactions between BOUNDARY *CLASS. The final[4] version of model 1 was:

normalised lag ~ class + boundary + class:boundary + (1 | speaker) + (1 | rword)

An interaction was found for CLASS*BOUNDARY, $F(2,285.623) = 3.425$, $p = 0.034$. Figure 6 is an interaction plot showing estimated means for CLASS*BOUNDARY. The broken line in Figure 6 shows the point where *rmax* is co-temporal with either offset of voicing, or onset of a following consonant. If data points occur above this line, *rmax* occurs after voicing offset, or during a following consonant (i.e. where consonants are nonlingual/velar). It can be seen that *rmax* occurs before voicing offset, or following consonant onset, in almost all boundary contexts for both social classes, with the exception of working-class /r/ in utterance-final position. This is a key finding given that utterance-final context is where /r/-lessness and /r/-weakening in Scottish English working-class speech have been most frequently reported in sociolinguistic studies through the years, see §1.1. The finding suggests again that a delay in the production of the anterior lingual /r/ gesture is responsible for working-class /r/ weakening in utterance-final contexts, due to auditory masking of part of the rhotic gesture, when vocal fold vibration ends before the lingual articulation is complete.

[4] Full model before step() comparison: normalised_lag ~ sex + class + stress + boundary + tauto_cons + class*boundary + (1|speaker)+ (1|rword)

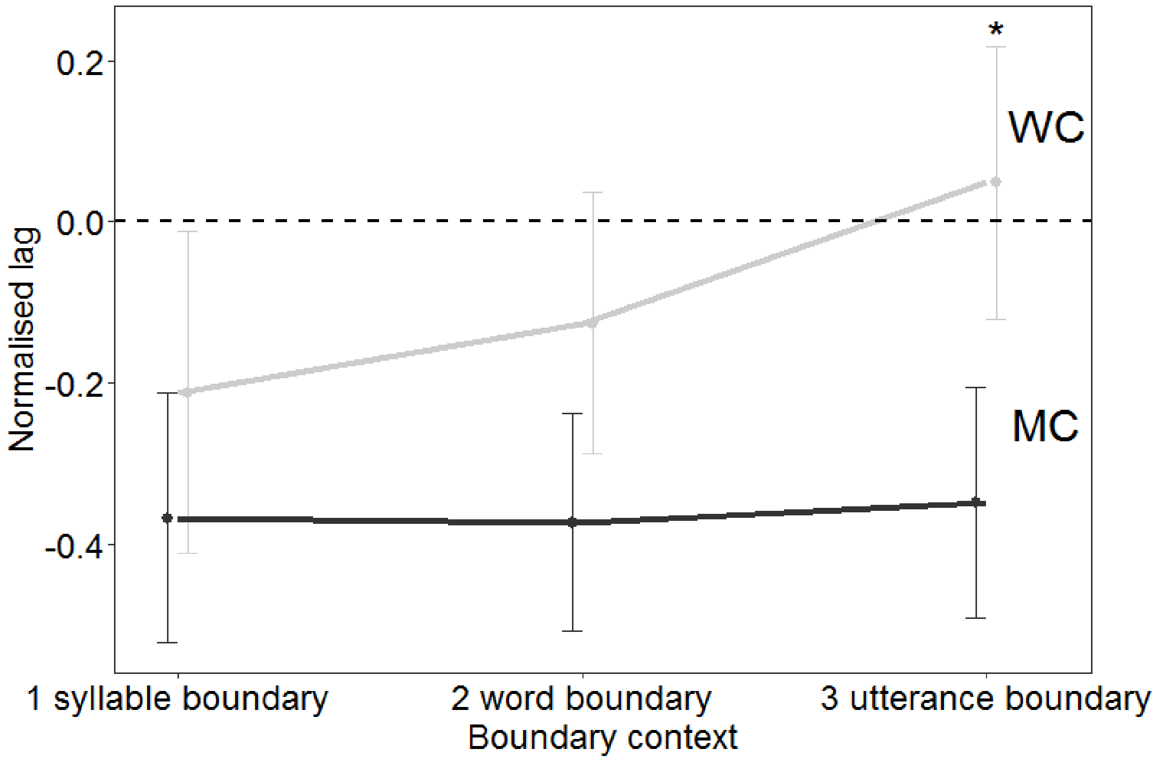


Figure 6: Interaction plot showing estimated means for *normalised lag* by BOUNDARY context and social CLASS, N=311.

Comparison of Figure 6 with Figure 5 (Rime duration by BOUNDARY context and social CLASS) shows an interesting class-based divergence in rhotic timing behaviour. The timing of the working-class anterior /r/ gesture is more delayed when the rime is lengthened, while the timing of the middle-class anterior /r/ gesture is unaffected by rime lengthening. This class-based difference in gesture-timing behaviour might relate to socially-stratified variation in /r/ tongue shape in this speech community (working-class - tip up, and middle-class - bunched). As hypothesised in §1.2.1., /r/ variants that involve formation of a constriction using the tongue body are more likely to have maxima close to the syllable nucleus, whereas /r/ variants that involve formation of a constriction with the tongue tip have more scope for anterior gesture drift towards (and beyond) syllable margins. This type of tongue shape variation might also explain why *normalised lag* is greater in all boundary contexts for working-class speakers than for middle-class speakers, see Figure 6.

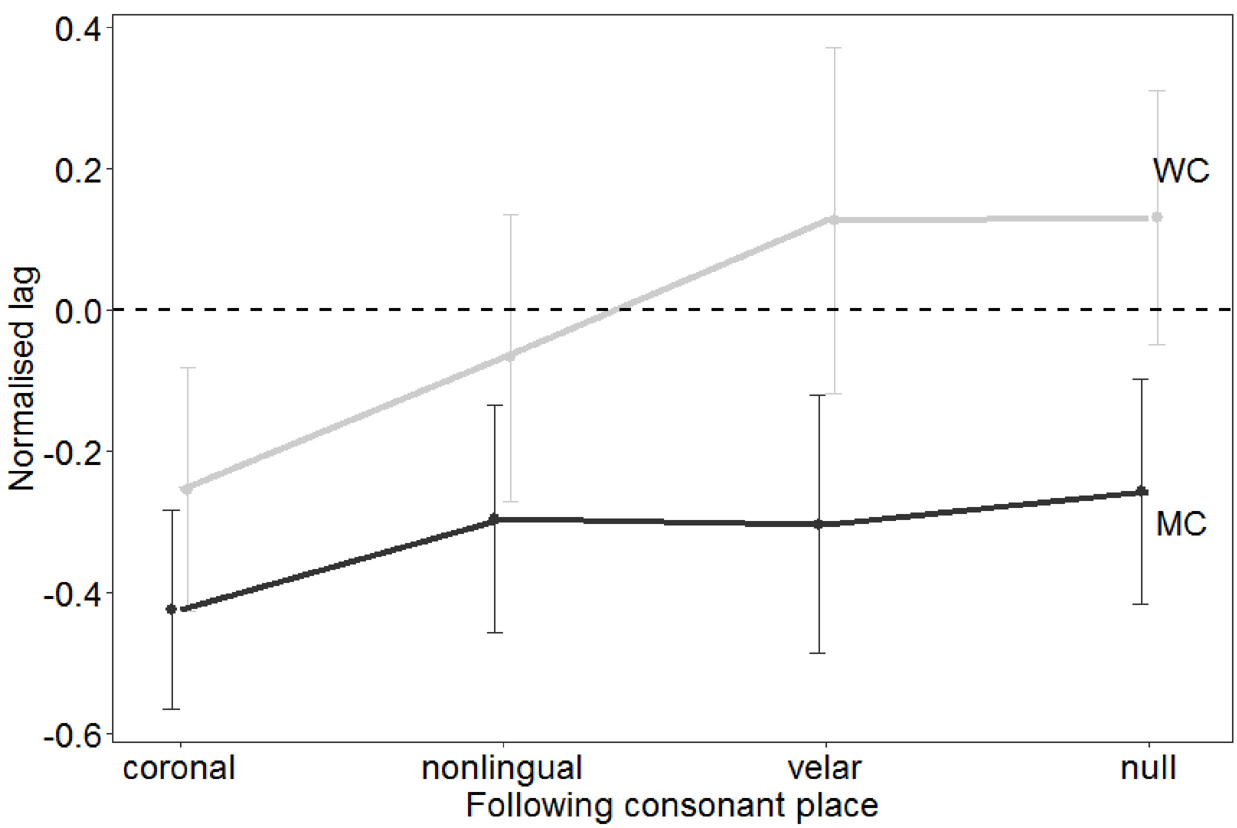
For both social classes, there were no significant differences in *normalised lag* across BOUNDARY contexts.

3.2.2.2 Normalised lag – model 2

The second model tested for the effects of SEX, CLASS, STRESS, and following consonant PLACE of articulation (including the level *null*, where /r/ was utterance final); testing for interactions between PLACE*CLASS. The final[5] version of model 2 was:

normalised lag ~ class + place + class:place + (1 | speaker) + (1 | rword)

An interaction was found between CLASS*PLACE $F(3,290.492) = 3.4524$, $p = 0.017$; Figure 7 shows estimated means for CLASS*PLACE. The broken line in Figure 7 again shows the point where *rmax* is co-temporal with either offset of voicing, or onset of a following consonant.



[5] Full model before step() comparison: normalized_lag ~ sex + class + stress + place + class*place + (1|speaker) + (1|rword)

Figure 7: Estimated means with standard errors for *normalised lag* by following consonant PLACE and social CLASS. N=311.

There were no significant differences between the two social classes for any of the levels of following-consonant PLACE, though the *null* level was approaching significance $t(19.2) = 3.352$, $p = 0.0534$. For both social-class groups, following consonants whose production was *coronal*, involving constriction formation with the anterior portion of the tongue (dental, alveolar and post-alveolar) conditioned the earliest preceding *rmax* gestures, while the latest *rmax* gestures occurred when no consonant followed (i.e. /r/ was utterance final). This means that the anterior /r/ gesture is effectively pushed backwards in time by a following consonant with a coronal specification, but is free to drift forward in time when no consonant follows. *Normalised lag* duration increased in a slightly different order for working-class and middle-class speakers:

(WC) *coronal* < *nonlingual* < *velar* < *null*
(MC) *coronal* < *velar* < *nonlingual* < *null*.

Again, it is possible that these subtle class-based differences in lingual gesture timing can be attributed to class-based variation in /r/ tongue shape; middle-class speakers may have earlier *rmax* gestures before velar consonants, because their bunched /r/s often involve constriction formation involving the tongue dorsum.

Again, for all levels of following-consonant PLACE, working-class speakers had more delayed estimated mean *normalised lag* than middle-class speakers, and estimated means for the *velar* and *null* places of articulation indicate that *rmax*, for this social-class group, occurred during the articulation of the velar consonant, and after voicing offset respectively. There is therefore, again, the potential for audible masking of working-class /r/ when followed by a velar consonant, or when it is in utterance-final position.

Across places of articulation for the following consonant, there were significant differences for estimated mean *normalised lag* between *coronal* and *velar* and *coronal* and *null* places of articulation for working-class speakers only, see Table 5.

Table 5: Other significant variation between FOLLOWING CONSONANT PLACE and social CLASS.

(WC) coronal – null ***	$t(273.9) = 5.335$ $p < 0.0001$
(WC) coronal – velar *	$t(235.1) = 3.515$ $p < 0.0122$

3.2.2.3 Normalised lag model comparison

An AIC-based comparison of the two models showed that model 2, containing the factor *following-consonant PLACE of articulation*, had the most parsimonious fit; model 1, AIC=186; model 2, AIC=175.

3.2.3 Audio r-index score

All coda-/r/ tokens were anonymously and randomly rated by the first author on a 5-point scale of rhoticity strength, from 1 = no audible /r/ to 5 = rhoticised vowel. The 10 tokens where an /r/ gesture was found to be absent in the articulatory-analysis phase were also rated in the auditory analysis. Their mean *r-index score* was 1.6 (midway between the "no /r/" and "derhoticised /r/" categories). The fact that some tokens with no /r/ articulation were identified as "derhoticised" is most likely due to the rater attending to cues associated with derhoticisation, such as pharyngealisation of the pre-rhotic vowel, which could be present even if no anterior /r/ gesture was present. These tokens with no anterior /r/ gesture were not included in the statistical analysis below.

Again, two models were run, substituting boundary CONTEXT and TAUTOSYLLABIC_CONSONANT in model 1, with following-consonant PLACE in model 2 to avoid collinearity.

3.2.3.1 Audio r-index score – model 1

The first model tested for the effects SEX, CLASS, STRESS, BOUNDARY context, and TAUTOSYLLABIC_CONSONANT; testing for interactions between BOUNDARY*CLASS. The final[6] version of model 1 was:

r_index ~ sex + class + boundary + class:boundary + (1 | speaker) + (1 | rword)

The fixed factor SEX was found to be significant $F(1, 11.19) = 10.0626$, $p = 0.0087$. Figure 8 shows that females had an estimated mean *r-index score* close to "retroflex"; 3.63 (+/- 0.217), just under one point higher than males' estimated mean *r-index score* 2.66 (+/- 0.227), which is close to "alveolar". This is not unexpected, as it is a common finding in sociolinguistic studies that females use more overtly prestigious

[6] Full model before step() comparison: r_index ~ sex + class + stress + boundary + tauto_cons + class*boundary + (1|speaker) + (1|rword)

variants than males (Labov, 2001) and strong rhoticity is overtly prestigious in the Central Scottish English speech community.

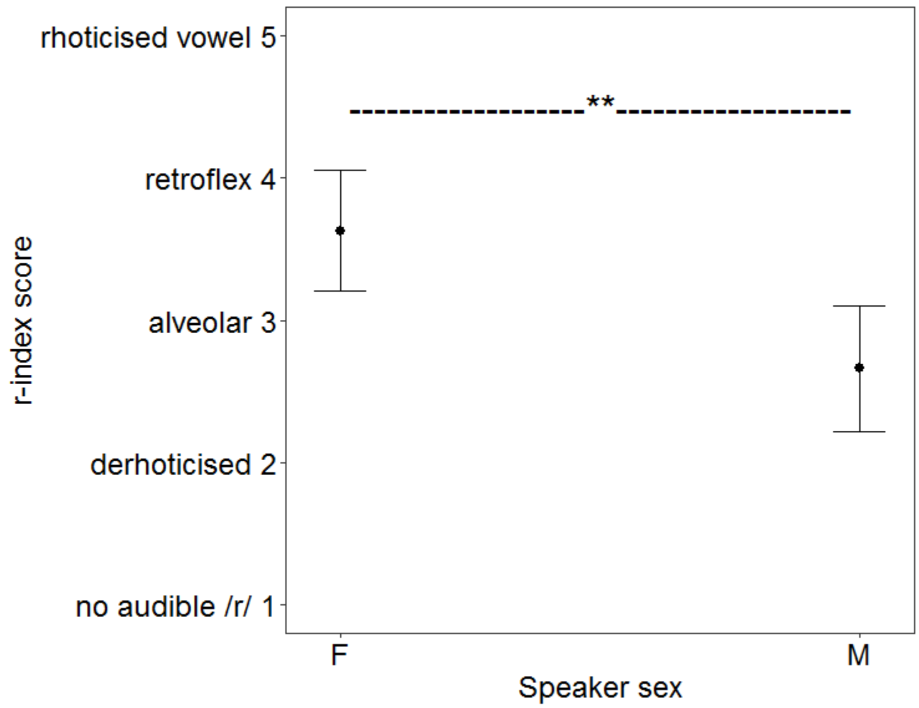


Figure 8: Estimated means with standard errors for *r-index score* by speaker SEX. N=311.

An interaction was also found between CLASS*CONTEXT $F(2,287.234) = 3.647$, $p = 0.0273$; significant differences were found between the two social classes for *word boundary* context $t(14) = 3.599$, $p = 0.0281$, and *utterance boundary* context $t(15.9) = 4.50$, $p=0.0041$, see Figure 9. Across all boundary contexts, middle-class speakers had estimated mean *r-index scores* that were greater (more strongly rhotic) than those of working-class speakers, but especially in *utterance-final* position, see Figure 9.

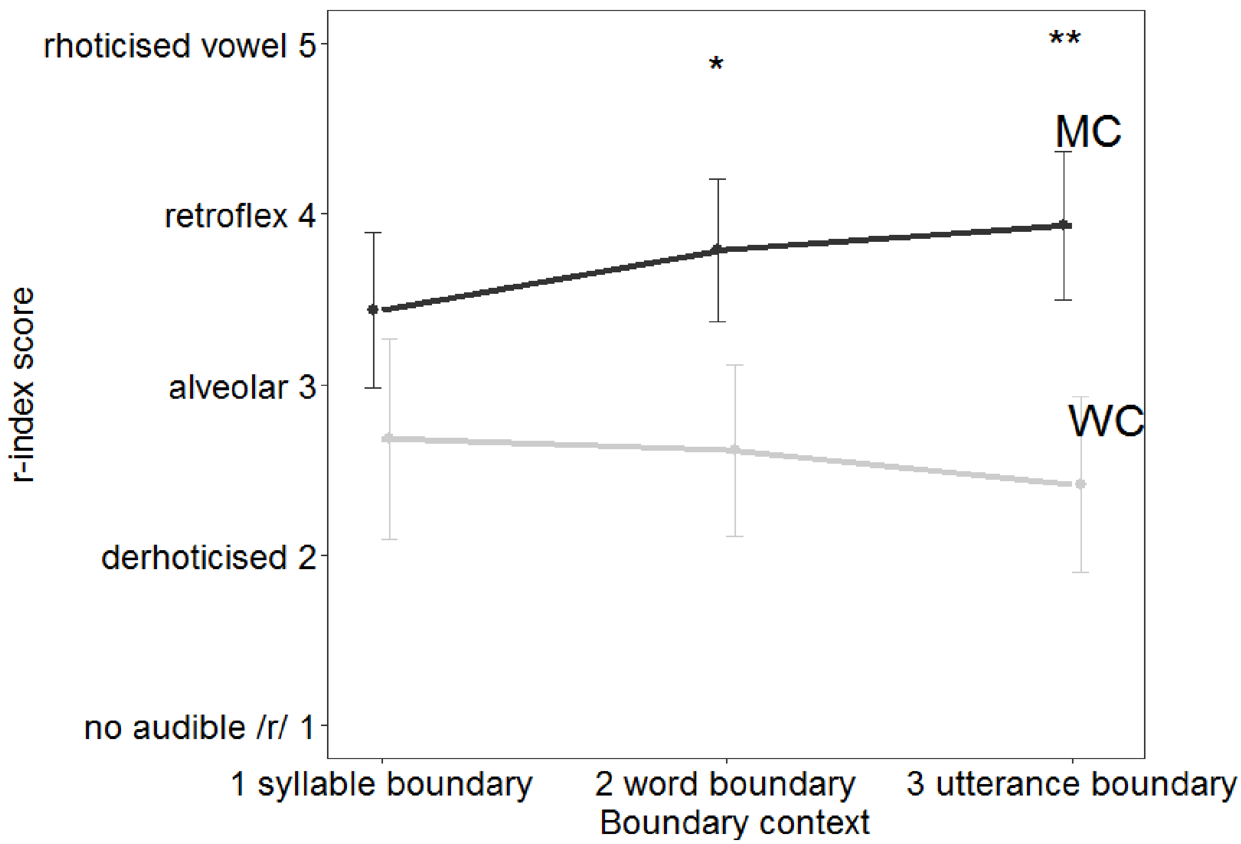


Figure 9: Interaction plot showing estimated means for *r-index score* by BOUNDARY context and social CLASS, N=311.

Middle-class speakers' estimated mean *r-index score* in utterance-final boundary context was 3.93, close to the "retroflex" category, and one and a half index points higher than working-class speakers' estimated mean *r-index score*, 2.41, between the "derhoticised" and "alveolar" categories. Within working-class and middle-class groups, there were no significant differences in *r-index scores* across boundary contexts. These results show a class-based division in /r/ strength that is accentuated in utterance-final position. Effectively, we see lenition of working-class utterance-final /r/, and fortition of middle-class utterance-final /r/. Working-class lenition in utterance-final position can be linked to anterior rhotic gesture delay, as gesture-timing results presented in §3.2.2 show that audible masking of the /r/ articulation could occur in utterance-final position due to the anterior lingual rhotic gesture being delayed beyond the point of voicing offset; however, fortition of middle-class /r/ in utterance final position is not clearly linked to gesture timing. Some other articulatory factor is causing utterance-final fortition of middle-class /r/.

3.2.3.2 Audio r-index score – model 2

The second model tested for the effects of SEX, CLASS, STRESS, and following consonant PLACE of articulation, testing for interactions between PLACE*CLASS. The final[7] version of model 2 was:

r_index ~ sex + class + place + (1 | speaker)

As with model 1, the final second model, included the fixed factor SEX, $F(1,11.108) = 8.018$, $p = 0.0161$; female speakers' estimated mean *r-index score*, 3.52 (between "alveolar" and "retroflex"), was around 1 point higher than male speakers' estimated mean *r-index score*, 2.51 (between "derhoticised" and "alveolar").

The final second model also included social CLASS as a fixed factor, $F(1,11.227) = 14.014$, $p = 0.0031$; estimated mean *r-index* scores for the levels of CLASS were MC=3.7 (close to the "retroflex" category), and 1.3 points higher than WC=2.4 (between the "derhoticised" and "alveolar" categories).

Finally, fixed factor following-consonant PLACE, $F(3,297.339) = 4.632$, $p = 0.0035$ was found to have a significant effect on *r-index score*. Figure 10 shows the impact of following consonant PLACE on the *r-index score*.

[7] Full model before step() comparison: r_index ~ sex + class + stress + place + class*place + (1|speaker)+ (1|rword)

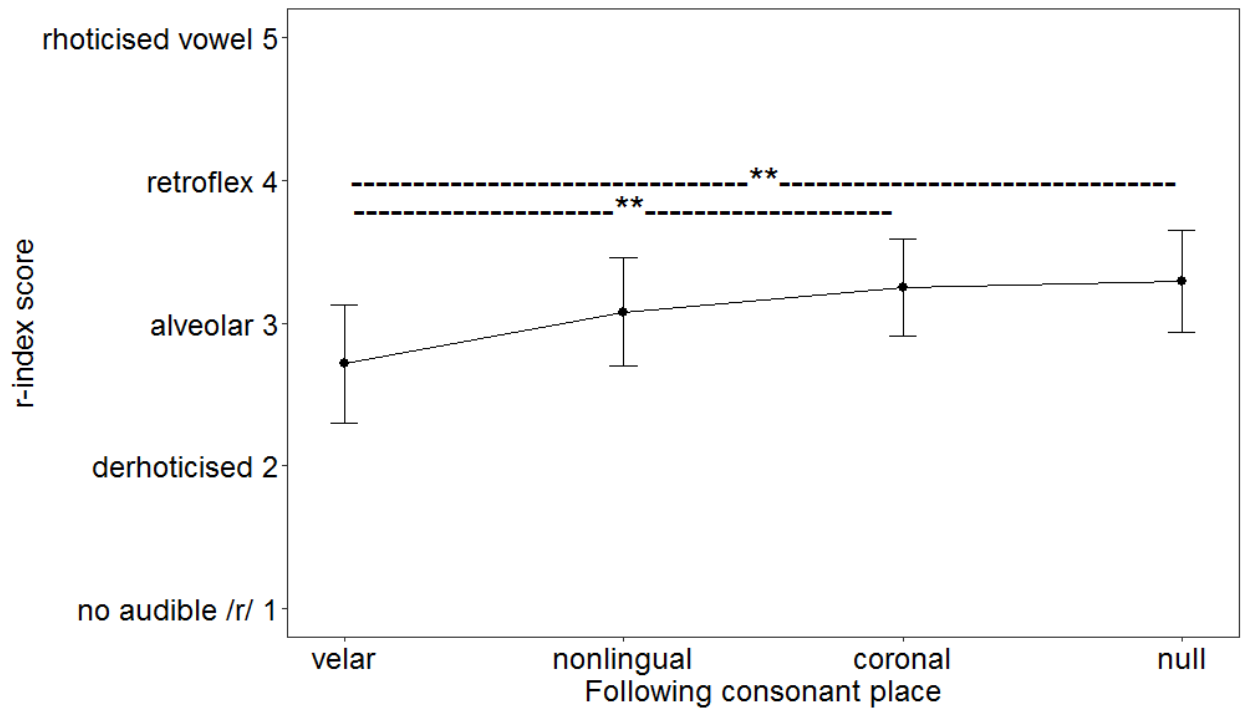


Figure 10: Estimated means with standard errors for *r-index score* by following consonant PLACE of articulation. N=311.

There were significant differences in *r-index score* between the velar place and coronal place of articulation of the following consonant, $t(296) = 3.438$, $p = 0.0037$; and also between velar and null $t(299) = 3.283$, $p = 0.0063$.

It might be expected that Figure 10 would mirror Figure 7 (showing timing-based variation across PLACE levels); that the following-consonant PLACE level conditioning the smallest degree of anterior gesture delay in /r/ (*coronal*), should condition the strongest percept of rhoticity, and conversely the PLACE level that conditioned the greatest degree of anterior gesture delay in /r/ (*null*), should condition the weakest percept of rhoticity. Figure 10 shows that in actual fact, these two PLACE levels condition similar estimated mean r-index scores; *coronal* (3.25) and *null* (3.29). In the absence of an interaction between CLASS and PLACE in this model for *r-index*, it is likely that some class-based variation in the *r-index scores* associated with the *null* level of PLACE is conflated in the current model.

3.2.3.3 r-index model comparison

An AIC-based comparison of the two r-index models showed that model 1, containing the factor BOUNDARY context, showed the most parsimonious fit, though marginally so; model 1, AIC=826; model 2, AIC=827.

3.3 Correlational analyses: normalised lag and r-index

Given the class-based variation observed for both *normalised lag* and *r-index score* in previous sections, separate Spearman's rank correlation tests, were run for working- and middle-class tokens for these two dependent variables using the cor.test() function in the "stats" R package (Core Team, 2018).

For working-class tokens, a significant negative correlation was identified, r_S= -0.36, p<0.001; the greater the anterior gesture lag for postvocalic /r/, the weaker its audible rhoticity, confirming the pattern previously found for Scottish English word-list recordings (Lawson et al, 2018). 80% of working-class tokens, see Figure 11 (left), are in the lower half of the r-index; categories 1-3 (no /r/, derhoticised, alveolar). Figure 11 (left) also shows that tokens rated (1) "no /r/" and (2) "derhoticised" tend to have anterior /r/ gesture maxima that occur after voicing offset, or following consonant onset (data points located to the right of the dotted line). However, for the tweleve working-class /r/ tokens that were rated 5 (rhoticised vowel), two had an /r/ gesture maximum after voice offset/following-consonant onset. This correlation test finding confirms that, in working-class spontaneous speech, lenition of /r/ is related to gesture delay, resulting in audible masking of some or all of the /r/ articulation.

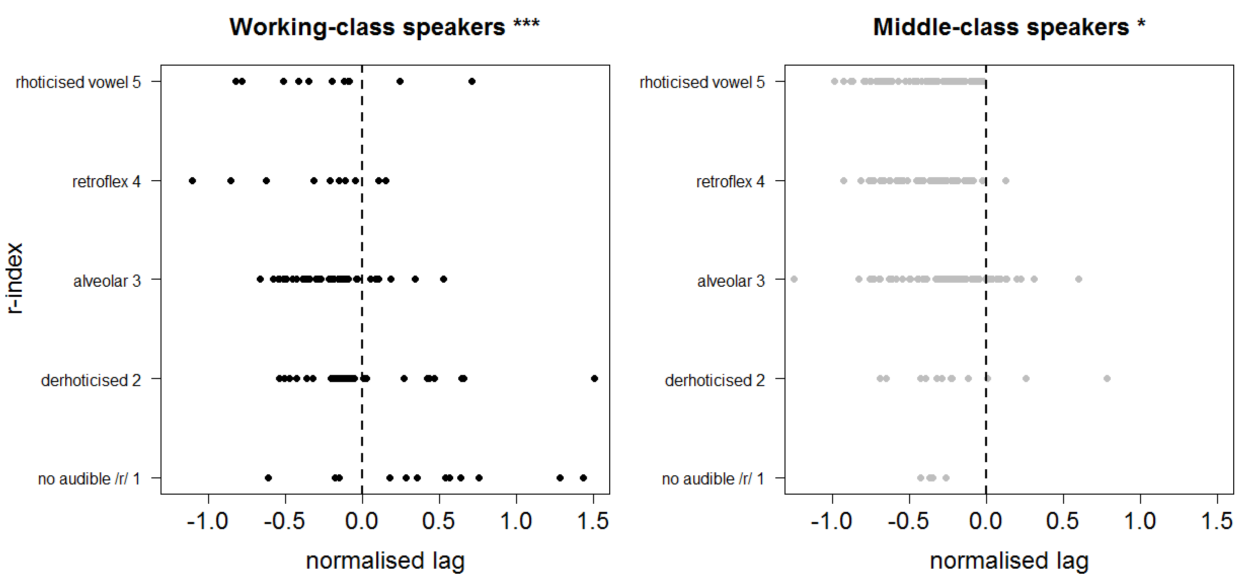


Figure 11: Correlation plots of *normalised lag* and *r-index score:* (left) working-class tokens, N=112 (right) middle class tokens, N=199. Vertical broken line indicates the time-point where the maximum of the anterior /r/ gesture is synchronous with voice offset, or following-consonant onset.

For middle-class tokens, there was a smaller significant negative correlation, r_S= -0.17, p=0.0188, showing that for middle-class speakers too, greater gesture lag corresponds to weaker-sounding /r/s. However, a weaker sounding /r/ for the middle-class group is more likely to be one classified as "alveolar" (N=69) than as "no /r/" (N=4), or "derhoticised" (N=13). The majority of tokens produced by middle-class speakers (91%), were in the upper half of the r-index; categories 3-5 (alveolar, retroflex, rhoticised vowel), see Figure 11, right. Tokens with anterior gesture maxima for /r/ occurring after voice offset, or following consonant onset (data points right of the dotted line in Figure 11 (right)) were rarer in the middle-class (8%), than in the working-class (28%) portion of the dataset.

4.0 Discussion

The current study investigates the phenomenon of increased /r/ weakening in utterance-final position in Scottish English and aimed to answer the following research questions:

(1) What impact does utterance-final position have on the timing and auditory quality of postvocalic /r/ in Scottish English?
(2) How does utterance position interact with other linguistic factors and with social factors in the /r/ weakening process?

We considered the effect of linguistic factors such as utterance position, syllable-stress, the presence of a following tautosyllabic consonant, and following consonant place of articulation, on three dependent variables: (i) *rime duration* (ii) *normalised anterior lingual gesture lag* during /r/ production and (iii) *strength of audible rhoticity*, making use of a socially-stratified audio-UTI conversational-speech dataset. This study assessed the impact of both social and linguistic factors on these dependent variables and in doing so, it provides a more comprehensive understanding of the processes involved in derhoticisation, and a better understanding of how prosody, phonotactics, speech articulation and social factors interact to bring about sound change.

Our study confirmed the hierarchical impact of linguistic boundaries on syllable *rime duration*. *rime duration* increased in the order: syllable boundary < word boundary < utterance boundary, and *rime duration* was significantly longer in utterance-boundary context than in syllable, or word boundary contexts. This finding agrees with findings of previous studies on the impact on boundary strength on rime duration and the duration of rime constituents, e.g. (Sproat, Fujimura 1993, Wightman, Shattuck-Hufnagel et al. 1992). However, it was additionally identified that there were significant differences in utterance-final rime duration according to social class; middle-class speakers had longer utterance-final rime duration than working-class speakers.

Analysis of *normalised lag* showed that working-class speakers had longer lags (more delayed anterior gesture maxima for /r/) across all boundary contexts, compared to middle-class speakers. In utterance-final position only, working-class anterior /r/ gesture maxima occurred, on average, after voicing offset, suggesting that lenition of /r/ in this boundary context, in the working-class social group, could indeed be due to audible masking of the /r/ articulation. Comparison of *rime duration* and *normalised lag* across boundary contexts also revealed interesting class-based variation; as rime durations increased across the three boundary conditions, so working-class /r/ anterior gesture maxima became more increasingly more delayed. However, this was not the case for middle-class speakers, where anterior /r/ gesture maximum timing appeared to be unaffected by *rime duration*. It is possible that another socially-stratified articulatory feature of Scottish English /r/ is responsible for this difference in rhotic timing behaviour. Working-class speakers in Central Scotland have previously been found to use tip-up/retroflex coda /r/ variants, while middle-class speakers use bunched coda /r/ variants (Lawson et al, 2011, Lawson et al, 2014a,). While tip-up /r/ involves forming a constriction using the tongue tip/blade at the postalveolar region, bunched /r/ involves a palatal constriction using the tongue dorsum or front. The tip/blade gesture hypothetically has greater freedom to drift to the margins of the audible syllable, and beyond, than the dorsal gesture of a bunched /r/, as the tip/blade gesture can act independently of the tongue body. It would seem that while bunched /r/ anterior maxima tend to occur temporally closer to the syllable nucleus, in some cases merging with the nuclear vowel to form a rhoticised vowel (Lawson et al, 2013), retroflex/tip-up maxima tend to occur closer to syllable margins and sometimes move beyond the audible margins of the syllable, i.e. occurring fully or partially after the offset of voicing.

An analysis of the impact of the post-rhotic consonant's place of articulation on the timing of the anterior lingual /r/ gesture showed that post-rhotic consonants with coronal places of articulation condition the earliest /r/ gestures, blocking the ability of anterior rhotic gestures to drift forward in time, while post-rhotic consonants with nonlingual, or velar, places of articulation allow the anterior lingual gesture of the rhotic to drift forward in time. The context that permitted the greatest temporal drift of the anterior /r/ gesture was the *null* context, i.e. no following consonant.

An audio analysis of rhotic strength identified that the context in which /r/ sounded weakest was the utterance boundary context, but only for working-class speakers. For middle-class speakers, utterance boundary context was associated with the *strongest* rhoticity ratings. In other words, we see utterance-final lenition of /r/ for working-class speakers, but fortition of /r/ in the same context for middle-class speakers. Both social-class groups showed a significant negative correlation between *normalised lag* and *r-index score* – longer lags are associated with weaker audible rhoticity, but the correlation was much stronger for the working-class group. Almost all of the /r/s produced by middle-class speakers occupied the top three (strongest) categories of the r-index.

While lenition of working-class utterance-final /r/ can demonstrably be linked to anterior lingual gesture delay, fortition of /r/ in middle-class speech is not related to gesture timing and therefore must be occurring by some other articulatory means. Again, the most likely candidates contributing to middle-class fortition of /r/ are (1) variation in tongue shape; middle-class use of bunched /r/ variants, which are associated with high F2 and low F3 values and a strong percept of rhoticity (Lawson et al, 2018), and (2) gesture magnitude; a narrowing of the root and/or palatal stricture of bunched /r/ to increase F2 and lower F3, see (MacFarlane and Stuart-Smith, 2012). It is unclear whether utterance-final strengthening of /r/ seen in this dataset is a common feature of middle-class Scottish English, as it has not been noted before, or whether it is a form of linguistic divergence resulting from interaction with working-class interlocutors.

Overall, therefore, it can be seen that utterance-final context conditions lengthening of syllable rimes in spontaneous speech, which also conditions greater anterior lingual gesture delay in working-class /r/. A lack of a following consonant and utterance-boundary context permits the anterior lingual gesture for /r/ to drift towards the margins of the syllable, and in some cases, beyond the margins of the acoustic syllable. Therefore, this study has identified an articulatory mechanism that can explain utterance-final lenition of /r/, a phenomenon that has been noted for over fifty years in Scottish sociolinguistic research (Romaine, 1978, Speitel and Johnston, 1983). However, this weakening mechanism does not appear to have the same effect on middle-class bunched /r/, which involves constriction formation with the tongue dorsum or front. Bunched anterior lingual articulations appear to be less free to drift towards the margins of the syllable. Additionally, we see audible fortition of middle-class /r/ in utterance-final position that is not linked to rhotic gesture timing. Another potential explanation for class-based timing variation in utterance-final position is that utterance-final position may be a key location for the performance of social-indexical phonetic variation; participants may be exploiting the salience of the location to emphasise /r/ variation and adjusting gesture timing, and perhaps also tongue shape and gesture magnitude, in order to achieve weaker or stronger /r/ variants in this boundary context.

5.0 Conclusion

In conclusion, current derhoticisation processes provide insight into how the historical /r/-loss process may have occurred, resulting in one of the most fundamental division in world Englishes: rhotic and nonrhotic varieties. By studying /r/ weakening in present-day Scottish English, we can see that /r/ loss may occur first in key locations, such as utterance-final position, and that prosodic factors, articulatory variation and social factors all play a role in enabling or blocking the derhoticisation sound change. The current study shows an interaction between linguistic and social factors, driving sound change in opposite directions. We see both lenition and fortition of postvocalic /r/ in utterance-final position in the same regional speech community depending on the social class of the speaker. Utterance-final position results in longer syllable rimes, and a lack of a following consonant allows the anterior /r/ gesture to drift to the edges of the acoustic syllable and beyond, making it partly, or completely, inaudible. However, this is only the case for working-class speakers in the study. For middle-class speakers, limited gesture drift occurs, despite the same conditions being present. We also see audible fortition of middle-class /r/ in utterance-final context, through means other than gesture timing. It is unclear whether socially-stratified variation in underlying tongue shape (tip up vs bunched) for syllable-coda /r/ entirely explains why utterance-final derhoticisation occurs in one class-based speech community and not the other, or to what extent we are also seeing speakers attempt to differentiate themselves from one another using /r/ strength. It is possible that the salience of utterance-final context makes it a key location for demonstrating the different forms of rhoticity, that are such an important indicator of social class in the speech community of Central Scotland. Further research should be undertaken in order to understand how rhotic tongue-shape variation constrains timing of tongue gestures.

Acknowledgements

This work was supported by the Economic and Social Research Council [grant numbers RES ES-N008189-1 and RES 062-23-3246];

Reference List

Boersma, P. and Weenink, D., 2013. *Praat: doing phonetics by computer.* 5.3.47 edn. http://www.praat.org/

Browman, C.P. and Goldstein, L., 1995. Gestural syllable position effects in American English. In: F. Bell-Berti and L.J. Raphael, eds, *Producing Speech: Contemporary Issues,* 19-33.

Core Team, R., 2018. *R: A Language and Environment for Statistical Computing.* Vienne, Austria: R Foundation for Statistical Computing.

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

Dobson, E.J., 1957. *English Pronunciation 1500 - 1700, volume II.* London: Oxford University Press.

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

Giles, S. and Moll, K., 1975. Cineflourographic study of selected allophones of English /l/. *Phonetica,* 31, 206-227.

Jauriberry, T., Sock, R., Hamm, A. and Pukli, M., 2012. Rhoticite et derhoticisation en anglais ecossais d'Ayrshire, *Proceedings of the Joint Conference JEP-TALN-RECITAL*, June 2012 2012, ATALA/AFCP, pp. 89-96.

Kochetov, A., 2006. Syllable position effects and gestural organization: Evidence from Russian. In: L. Goldstein, D.H. Whalen and C. Best, eds, *Papers in Laboratory Phonology VIII.* New York: Mouton de Gruyter, pp. 565-588.

Krivokapic, J. and Byrd, D., 2012. Prosodic boundary strength: An articulatory and perceptual study. *Journal of Phonetics,* 40(3), 430-442.

Kuznetsova, A., Brockhoff, P.B. and Christensen, R.H.B., 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software,* 82(13), 1-26.

Labov, W., 2001. *Principles of Linguistic Change, Volume 2: Social Factors.* Oxford: Blackwell.

Lawson, E., Leplatre, G., Stuart-Smith, J. and Scobbie, J.M., 2019. The effects of syllable and utterance position on tongue shape and gestural magnitude in /l/ and /r/. *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne. 4-10 August 2019* 2019, pp. 432-436.

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

Lawson, E., Scobbie, J.M. and Stuart-Smith, J., 2014a. A socio-articulatory study of Scottish rhoticity. In: R. Lawson, ed, *Sociolinguistics in Scotland.* London: Palgrave Macmillan, 53-78.

Lawson, E. and Stuart-Smith, J., 2019. The effects of syllable and sentential position on the timing of lingual gestures in /l/ and /r/, *Proceedings of the XIXth International Congress of Phonetic Sciences. 4-10th August, 2019.* 2019, pp. 547-551.

Lawson, E., Stuart-Smith, J. and Scobbie, J.M., 2018. The role of gesture delay in coda /r/ weakening: An articulatory, auditory and acoustic study. *Journal of the Acoustical Society of America,* 143(3), 1646-1657.

Lawson, E., Stuart-Smith, J. and Scobbie, J.M., 2014b. A mimicry study of adaptation towards socially-salient tongue shape variants, *Selected papers from NWAV 42*, October 17th - 20th 2013 2014, University of Pennsylvania Working Papers in Linguistics, pp. 101-110.

Lawson, E., Stuart-Smith, J. and Scobbie, J.M., 2008. Articulatory Insights into Language Variation and Change: Preliminary Findings from an Ultrasound Study of Derhoticization in Scottish English. *University of Pennsylvania Working Papers in Linguistics,* 14(2), article 13- http://repository.upenn.edu/pwpl/vol14/iss2/13.

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

Lehiste, I., 1973. Rhythmic units and syntactic units in production and perception. *Journal of the Acoustical Society of America,* 54(7), 1228-1234.

Lenth, R., 2019. *emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.3.4.* 1.3.4 edn.

Maguire, W., McMahon, A., Heggarty, P. and Dediu, D., 2010. The past, present, and future of English dialects: Quantifying convergence, divergence, and dynamic equilibrium. *Language Variation and Change,* 22(1), 69-104.

McAllister, A., 1938. *A Year's Course in Speech Training.* London: University of London Press Ltd.

McMahon, A., Heggarty, P., McMahon, R. and Maguire, W., 2007. The sound patterns of Englishes: representing phonetic similarity. *English Language and Linguistics,* 11(1), 113-142.

Mielke, J., Baker, A. and Archangeli, D., 2010. Jeff Mielke, Adam Baker, and Diana Archangeli. Variability and homogeneity in American English /r/ allophony and /s/ retraction. *Variation, Detail, and Representation. (LabPhon 10).* 2010, 699-719.

Price, P.J., Ostendorf, M., Shattuck-Hufnagel, S. and Fong, C., 1991. The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America,* 90(6), 2956-2970.

Primer, S., 1890. Pronunciation of Fredericksburg, V.A. *Publications of the Modern Language Association of America,* 5.

Recasens, D. and Farnetani, E., 1994. Spatiotemporal properties of different allophones of /l/: phonological implications, W.U. Dressler, M. Prinzhorn and J.R. Rennison, eds. In: *Phonologica 1992: Proceedings of the 7th International Phonology Meeting* 1994, Rosenberg & Sellier, pp. 195-204.

Romaine, S., 1978. Postvocalic /r/ in Scottish English: Sound change in progress. In: P. Trudgill, ed, *Sociolinguistic Patterns in British English.* pp. 144-158.

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, *Proceedings of the Eighth International Seminar on Speech Production (ISSP), Strasbourg*, 8-12 December, 2008 2008, pp. 373-376.

Speitel, H.H. and Johnston, P.A., 1983. *A Sociolinguistic Investigation of Edinburgh Speech.* End of Grant Report. Economic and Social Research Council.

Sproat, R. and Fujimura, O., 1993. Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics,* 21(3), 292-311.

Stuart-Smith, J., 2007. A sociophonetic investigation of postvocalic /r/ in Glaswegian adolescents, Trouvain, J., Barry, W.J., ed. In: *Proceedings of the 16th International Congress of Phonetic Sciences*, 6 - 10 August 2007 2007, Universitat des Saarlandes, pp. 1449-1452.

Stuart-Smith, J., 2003. The phonology of modern urban Scots. In: J. Corbett, J.D. McClure and J. Stuart-Smith, eds, *The Edinburgh Companion to Scots.* 1 edn. Edinburgh, U.K.: Edinburgh University Press, pp. 110-137.

Stuart-Smith, J. and Lawson, E., 2017. Scotland: Glasgow and the Central Belt. In: R. Hickey, ed, *Listening to the Past. Series: Studies in English Language.* CUP: Cambridge, .

Walker, J., 1791. *A Critical Pronouncing Dictionary.* London: G. G. J. & J. Robinson.

Wells, J., C., 1982a. *Accents of English 1: Introduction.* Cambridge: Cambridge University Press.

Wells, J., C., 1982b. *Accents of English 2: The British Isles.* Cambridge: Cambridge University Press.

Wightman, C.W., Shattuck-Hufnagel, S., Ostendorf, M. and Price, P.J., 1992. Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America,* 91(3), 1701-1717.

Wrench, A., 2012. *Articulate Assistant Advanced User Guide.* 2.14 edn. Edinburgh: Articulate Instruments Ltd.