View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Queen Margaret University eResearch

THE ROLE OF ANTERIOR LINGUAL GESTURE DELAY IN CODA /r/ LENITION: AN ULTRASOUND TONGUE IMAGING STUDY

Eleanor Lawson[1], James M. Scobbie [1] and Jane Stuart-Smith[2]

[1]CASL, Queen Margaret University Edinburgh, [2]GULP, University of Glasgow

elawson@qmu.ac.uk

ABSTRACT

We investigate the contribution that lingual gesture delay makes to lenition of postvocalic /r/. This study uses a socially-stratified, audio-ultrasound corpus of Scottish English containing recordings from two sociolects; one with postvocalic /r/ weakening and the other with strengthening. We quantify auditory strength of rhoticity and the timing of the anterior lingual gesture relative to the offset of voicing in CVr words: *bar, bore, fur*, or onset of a following consonant in CVrC words: *farm, herb, burp*, in order to show that there is a statistically significant correlation between weak rhoticity and a late articulatory gesture. Our ultrasound data also show that during the process of final consonant vocalization/deletion, underlying articulatory gestures may persist.

Keywords: rhoticity; ultrasound; sociophonetics; sound change; intergestural timing.

1. INTRODUCTION

The notion that there is a phonetic basis for the cross-linguistic tendency of coda consonants to lenite, vocalise or be deleted has, for a long time, been of interest to phoneticians and phonologists. Over the past few decades, articulatory studies of speech sounds have revealed the role that articulatory gesture timing plays in determining the phonetic quality of coda consonants. Sproat & Fujimura [12], Browman & Goldstein [2] and Krakow [4] have all highlighted the tendency for the multiple articulatory gestures that make up some consonants to be more synchronous when the consonant is in onset position and less synchronous when the consonant is in coda position. The sequence of gestures has also found to be determined by syllable position in many studies. Sproat and Fujimura's study of American English /l/, for example, used an articulatory analysis technique, x-ray microbeam, to show that there could be variable sequencing between the tongue dorsum retraction maximum and apical advancement maximum [12]. In onset position, the apical gesture for /l/ occurred before the dorsum retraction gesture, whereas the reverse was true in coda position. This variation in gestural sequencing was found to correlate with the percept of clear (more palatalized) and dark (more velarized or vocalized) /l/ variants in onset and coda position respectively.

The possibility that a late apical advancement gesture in syllable coda position might result in auditory deletion of a segment has been suggested by Recasens and Farnetani [9], who 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 vocalization at the acoustic level, if not articulatory level.

The present study uses ultrasound tongue imaging to investigate the role of lingual gesture timing in the lenition of postvocalic /r/ in a rhotic variety of English.

1.1. Scottish rhoticity

For several decades, researchers have noted lenition of coda /r/ in the English of Central Scotland [11], [12], [13], a variety of English that is usually described as rhotic. These mainly auditory-acoustic studies have shown that strength of rhoticity is socially stratified, with middle-class speakers preserving rhoticity, while working-class speakers produce greater quantities of weakly rhotic postvocalic /r/s. Minimal pairs such as *bud/bird* /bʌd/bʌrd/ and *cod/cord* /kɔd/kɔrd/ can sound very similar, but they can, for the most part, still be differentiated by local speakers [8], most likely due to qualitative adjustments (pharyngealisation or velarisation) in the prerhotic vowel; nevertheless, identification of a distinct rhotic segment can be difficult. There is often a great deal of inconsistency in how lenited /r/ variants are transcribed [13]. To date, there has been no systematic quantitative articulatory analysis of gesture timing in Scottish /r/. This paper presents such a gesture-timing study,

using an audio-ultrasound corpus of Glaswegian adolescent speech.

2. DATA AND METHOD

2.1. Participants

The Western Central Belt audio-ultrasound tongue imaging corpus (henceforth WCB12) was collected in 2012 in Glasgow. 16 adolescents aged 12-13 were recorded; four males and four females each, from two schools, one in an affluent area of the city and one in a socioeconomically deprived area of the city.

2.2 The ultrasound tongue imaging recordings

Informants were recorded with audio and ultrasound tongue imaging in an IAC sound-attenuated recording booth at the University of Glasgow. All noise-making equipment such as the ultrasound machine and PC were located outside of the recording booth.

To reduce pitch, roll and yaw of the ultrasound probe in relation to the speaker's cranium, an Articulate Instruments stabilising headset [10] was fitted to each speaker's head with the ultrasound probe held in place underneath the chin by the headset.

Single word prompts with no carrier (thus avoiding coarticulatory effects) were presented orthographically to participants on a monitor. Audio recordings were made using a Beyer-Dynamic Opus 55 headworn microphone. The recordings were sampled at 22kHz and a video-output Mindray DP2200 ultrasound machine, set to NTSC video format, created ultrasound video at a target rate of 29.97fps. The frame rate of the UTI video was doubled to circa 59.94fps by deinterlacing each video frame post hoc.

2.3. Word list

The word list had 45 monosyllabic items containing postvocalic /r/, 25 of which were CVr words, 5 of which were CVrC words and 15 of which were CVrC nonsense words. There were also 98 distractors, some of which were real words and some nonsense words. The inclusion of nonsense words does not directly relate to the design of the current study; another aim of the data collection was to obtain a set of stimuli for a subsequent nonsense-word mimicry experiment [6]. The /r/-ful nonsense words were included in the current study in order to increase the number of tokens of /r/.

For all words in the study, we avoided lingual consonants in order to restrict potential coarticulatory effects on /r/. This constraint again related to another study pertaining to tongue shape during /r/ production [5].

2.4. Audio-Video synchronisation

Both the audio channel and the video channel from the video-output ultrasound machine passed through a SynchBrightUp unit (created by Articulate Instruments) which superimposed a white square on the video signal simultaneously with a tone and pulses on the audio signal, near the beginning of each new recording. These signals were used by the analysis software [16] to re-establish the UTI video frame rate and to resynchronise audio and video on each recording.

The synchronisation then had to be adjusted to take into account the duration of the internal delay in image formation which video-output ultrasound scanners impose between completion of the scan cycles at the probe and the output of each NTSC video frame [14], [15]. Such adjustment has to be based on an empirical baseline measurement of the internal delay of each scanner; a "tap test" that measures the delay between the audio and visual record of a microphone capsule being tapped onto the ultrasound probe.

The DP2200 video-output machine used in this study has an average image processing delay of 20ms) between the acquisition of the ultrasound signal and the export of each video frame. In order to take account of this delay, a -20ms lag was introduced to the video signal during synchronisation of audio and video. The variable processing delay means that alignment of each video frame to the audio signal is not exact to the millisecond level; however variation in the amount of time it takes the DP2200 to create a video frame is random and slight inconsistencies in synchronization of video and audio would not account for any statistically significant patterns of timing variation between social-class groups in this study. In other words, the lack of fine-grained synchronization acts as general noise, affecting all the data, over which robust patterns of variation can emerge and be quantified.

2.5. Tongue gesture timing annotation

We studied the timing of the anterior lingual gesture for postvocalic /r/, and its relation to the offset of voicing in CVr words, or the onset of a following labial consonant in CVrC words. Four main temporal events were annotated for each /r/-ful token.

(1) *rmax* - the temporal location of the /r/'s maximal anterior constriction gesture. Annotated at the first video frame where the maximum constriction is achieved.
(2) *V-onset* - the temporal location of the onset of the vowel in CVr and CVrC words.
(3) *voice-offset* - the temporal location of the offset of voicing in CVr words.
(4) *C-onset* - the temporal location of the closure in the final labial consonant in CVrC words.

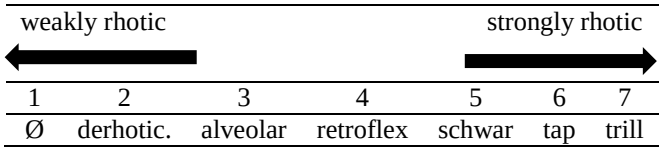
rmax was determined by visually inspecting the ultrasound tongue imaging video. The other three annotations were made based on acoustic information using Praat [1]. The durational difference between *rmax* and *voice-offset* and *rmax* the *C-onset* will both be called *lag*. A positive *lag* indicates that *rmax* occurs after *voice-offset* / *C-onset*, whichever is appropriate for the CVr or CVrC word concerned. A negative *lag* indicates that *rmax* occurs before these events. In order to account for variation in speech rate, *lag* was normalised by dividing it by the duration of the vowel + /r/ segment, which means that *normalised lag* is expressed as a proportion of the vowel-plus-/r/ (Vr) section of the syllable rime.

2.5. Auditory analysis - /r/ index

All tokens of /r/ were rated on a 7-point rhoticity index, see Fig. 1. However, only one male speaker (GWM1) was found to produce tapped and trilled variants. GWM1's tokens were excluded from further analysis, as he produced only strongly trilled tokens of /r/ throughout the word-list recording, but not in spontaneous speech.

Auditory classification was carried out by the authors using a Praat multiple-forced-choice experiment interface, which presented randomised anonymous audio recordings of single words to the classifiers, who matched each audio stimulus to the most appropriate phonetic category. Category labels registered in the MFC as numbers ranging from 1 (Ø = no /r/) to 7 (trill). Up to twenty replays of each stimulus were allowed before a choice of category had to be made. 441 tokens (after exclusion of GWM1's tokens) were classified over several rating sessions and an average rhoticity index value was calculated for each token.

Figure 1: 7-point rhoticity index.



3. RESULTS

3.1. Anterior lingual gesture timing

Figure 2 presents as boxplots *normalised lag* by sociogender group. The horizontal broken line in Figure 2 represents zero lag, i.e. the point at which *rmax* co-occurs with *voice-offset*/*C-onset*. Datapoints below this line indicate the anterior /r/ gesture occurred before the offset of voicing and is likely to have been completely audible. For datapoints above this line, some or all of the anterior lingual /r/ gesture will have occurred after the offset of voicing, or during the articulation of a following labial consonant and therefore will be masked to varying degrees.

Figure 2: Boxplots showing normalized gesture lag by sociogender group, N=411.

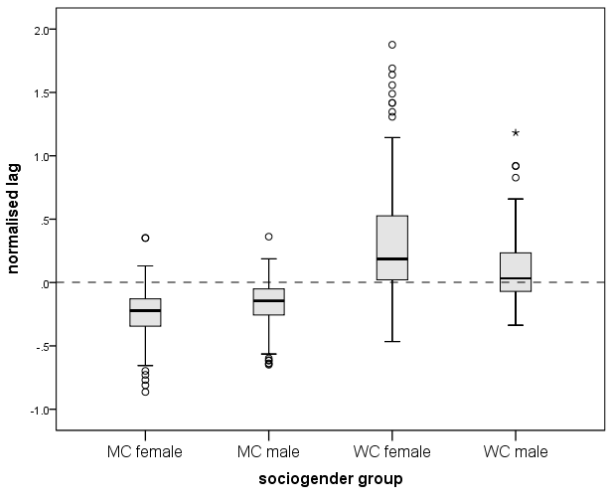


Figure 2 shows that the majority of the middle-class tokens (95% of the female middle-class tokens and 89% of the male middle-class tokens) had an anterior /r/ articulation that reached its maximum before the offset of voicing, or before the onset of the following labial consonant. Conversely, a large proportion of the working-class tokens contained an /r/ articulation that reached its maximum after the offset of voicing, or after the onset of the following labial consonant (78% of the female and 71% of the male working-class /r/ tokens). A oneway ANOVA with posthoc Bonferroni tests showed no significant. difference between the middle-class males and

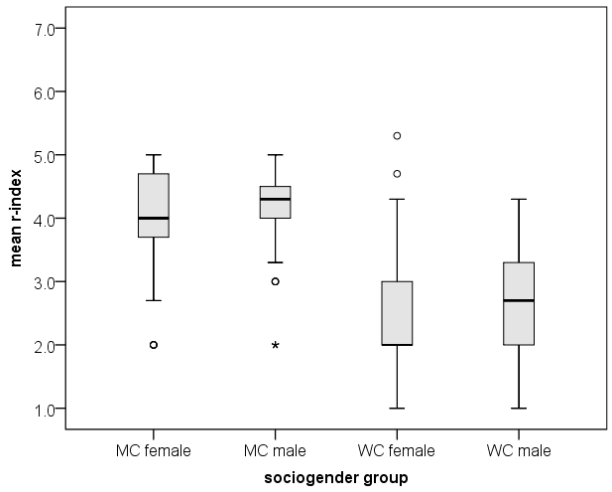
females' *normalised lag*, but significant differences to the $p<0.001$ level between the *normalised lag* of all other sociogender groups.

Among the working-class females, all of the positive outliers shown in Figure 2 were produced by one speaker, GWF1. Most of the negative outliers produced by middle-class speakers were instances of words where /r/ had merged with a preceding /ɪ/, /ɛ/, or /ʌ/ vowel, resulting in a monophthongal rhoticised vowel [ɚ], e.g. [fɚ] for *fur*, see [7].

3.2. Auditory /r/-index scores

Figure 3 below presents as boxplots the mean auditory /r/-index scores for tokens of /r/, organised by sociogender group. The higher the index score, the stronger the rhotic quality of the token.

Figure 3: Boxplots showing mean auditory /r/-index score by sociogender group, N=441.



Middle-class males and females had a mean /r/-index score of around 4, corresponding to the 'retroflex' category, while the working-class males and females had a mean score of around 2.5 (in between 'derhotic' and 'alveolar'). Figure 3 shows an inverse picture of Figure 2; the greater the *normalised lag*, the lower the /r/-index score. That is, the more delayed the anterior lingual gesture, the less rhotic the token sounded. A Spearman's correlation test found a significant negative correlation between *normalised lag* and mean /r/-index score $r_s=0.682$, $p<0.001$.

Outliers in the WC female group were *bar* produced by GWF4 and *far* GWF2, both of which were rated as strongly rhotic. Outliers in the MC female and male groups are mostly *er*, *ir* words; *verb*, *perm*, *firm*, produced by a range of speakers, but also four instances of the word *form*, which were rated as strongly rhotic.

Despite quite high levels of 'no /r/' classification at the auditory level for the working-class speaker group, almost every r-word token contained a lingual /r/ gesture at the articulatory level. Ultrasound video recordings associated with tokens that had been rated 'derhotic' or even 'no /r/' by one or more classifier, did in fact have /r/ gestures comparable (in terms of tongue configuration and degree of stricture) with those produced when postvocalic /r/ was completely audible. There were, a few instances of /r/ produced by speakers in the working-class cohort (male speaker GWM2 and female speaker GWF3) that showed articulatory reduction – a more subtle raising of the tongue during the production of /r/. These were always instances of words where /r/ followed a high vowel, e.g. *beer, peer, ear, fear, bear, air, pair, hair, oar, moor, boor, poor* and where an auditorily salient epenthetic glide vowel was present, usually [ʌ] or [ɐ], e.g. [mʉʌ] *moor*, [biɐ] *beer*. One token, produced by GWF3, was tentatively labeled as articulatorily /r/-less; however, even this token, *boor*, showed evidence of subtle fronting and raising of the tongue front, although with no subsequent relaxing of the tongue posture to a rest position.

4. DISCUSSION

This study shows that even with data recorded using portable video-output ultrasound machines, insights into temporal gestural organisation can be gained. Our study also shows that the gestural dissociation associated with coda liquids is contributing to the weakening of auditory percept of rhoticity found in working-class Scottish English. The fact that /r/ is present at the articulatory level, but that part or all of the anterior /r/ gesture is masked either by lack of voicing or by a following labial consonant helps to explain some of the difficulty previous researchers have found in coding /r/ tokens in studies of working-class Scottish speech, [13]. Articulatory reduction was also found to be present in our articulatory data, though it was associated with particular speakers and not necessarily those who produced the greatest proportion of weakly rhotic variants. Our study shows that both internal factors such as gestural dissociation in syllable-coda position and social-indexical factors such as social class, and perhaps also gender, are driving the weakening of rhoticity in Scottish English.

5. REFERENCES

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

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

[3] Jauriberry, T., Sock, R., Hamm, A. and Pukli, M. eds. 2012. Rhoticite et derhoticisation en anglais ecossais d'Ayrshire. *Proceedings of the Joint Conference JEP-TALN-RECITAL.*, June 2012, Grenoble, France: ATALA/AFCP.

[4] Krakow, R.A. 1999. Physiological organization of syllables: a review. *J. Phon.*, 27 (1) 1, pp.23-54.

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

[6] 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,Pittsburgh, USA,Pennsylvania: University of Pennsylvania Working Papers in Linguistics.

[7] 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 (3–4) 0, pp.198-210.

[8] Lennon, R. 2013. The effect of experience in cross-dialect perception: Parsing /r/ in Glaswegian. *Unpublished dissertation Submitted for the degree of Master of Science in English Language and Linguistics in the School of Critical Studies. University of Glasgow.*

[9] Recasens, D. and Farnetani, E. 1994. Spatiotemporal properties of different allophones of /l/: phonological implications. *Phonologica 1992: Proceedings of the 7th International Phonology Meeting.* Torino: Rosenberg & Sellier.

[10] Scobbie, J. M., Wrench, A. A. and van der Linden, Marietta eds. 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.

[11] Speitel, H.H. and Johnston, P.A., 1983. A Sociolinguistic Investigation of Edinburgh Speech. *Final report submitted to the Economic and Social Research Council.*

[12] Sproat, R. and Fujimura, O. 1993. Allophonic variation in English /l/ and its implications for phonetic implementation. *J. Phon.*, 21, 292-311.

[13] Stuart-Smith, J. ed. 2007. A sociophonetic investigation of postvocalic /r/ in Glaswegian adolescents. *Proceedings of the 16th International Congress of Phonetic Sciences.* , 6 - 10 August 2007,Saarbrucken, Germany: Universitat des Saarlandes.

[14] Wrench, A. A. and Scobbie, J. M. 2006. Spatio-temporal inaccuracies of video-based ultrasound images of the tongue. Proceedings of the 7th International Seminar on Speech Production. pp. 451-458.

[15] Wrench, A. A. and Scobbie, J. M., 2008. High-speed Cineloop Ultrasound vs. Video Ultrasound Tongue Imaging: Comparison of Front and Back Lingual Gesture Location and Relative Timing. Proceedings of the Eighth International Seminar on Speech Production (ISSP).

[16] Wrench, A. 2012. Articulate Assistant Advanced User Guide. 2.14th ed. Edinburgh: Articulate Instruments Ltd.

6. ACKNOWLEDGEMENTS

This study was funded by the UK Economic & Social Research Council (RES-062-23-3246).