APPARENT TIME CHANGE IN THE ARTICULATION OF ONSET RHOTICS IN SOUTHERN BRITISH ENGLISH

Patrycja Strycharczuk[1], Susan Lloyd[2] & James M. Scobbie[2]

[1]University of Manchester; [2]Queen Margaret University, Edinburgh
patrycja.strycharczuk@manchester.ac.uk; SLloyd@qmu.ac.uk; JScobbie@qmu.ac.uk

ABSTRACT

The /r/ phoneme in Anglo-English is known to correspond to a number of relatively distinct articulatory variants. However, little is known about the social structure of this variation. In this study, we investigate the effect of two social factors, age and gender, on the production of /r/, in a sample of 36 speakers from the South of England. We analysed ultrasound images of pre-vocalic /r/ tokens. We measured the distances between the short tendon and 11 points on the tongue surface. We compared these distances across speakers in representative /r/ frames. We find an apparent time difference whereby the distance between the tongue tip and the short tendon reduced in apparent time, potentially signalling an ongoing sound change from a tip-up to tip-down /r/.

Keywords: rhotics; variation; Southern British English; ultrasound; change

1. INTRODUCTION

Southern British English (SBE) rhotics are well-studied, but mostly from the point of view of whether they occur or not in specific contexts. In contrast, the quality of the /r/ in this variety has received less empirical attention. It is commonly described as a post-alveolar approximant [1, 2], although other variants, notably labio-dental, have also been documented [3, 4, 5].

An early articulatory study of /r/ in Anglo-English shows pronunciation by two speakers to be a retroflex approximant [ɻ] in onset position, and a vowel otherwise [6]. A recent study combining ultrasound and a video of the lips documents the articulation of /r/ in a geographically diverse sample of 24 speakers from England [7]. This study captures considerable articulatory variation, spanning three variants of tip-up /r/, two of which are types of voiced post-alveolar approximant [ɹ], while the other is a retroflex approximant [ɻ], as well as two types of tip-down bunched /r/. These variants resemble those found in other varieties of English, including Scottish English [8, 9], American English [10], and New Zealand English [11]. Bunched /r/ does not appear in most earlier descriptions of Anglo-English, even as an coarticulatory variant in supportive environments. As far as constraints on rhotic variation are concerned, tip-up /r/ is generally less likely to occur preceding front vowels than preceding back vowels. Otherwise, the variation is not known to be systematically structured.

In this study, we explore whether rhotic variation in SBE, a dialect of Anglo-Eglish, is systematically conditioned by social factors, specifically age and gender. Investigating these factors from a variationist point of view requires a relatively large sample, which can be challenging in the context of an articulatory study. However, articulatory data can be crucial in case of rhotics, given acoustic similarities between multiple articulatory /r/ variants. In this project, we used data collected at a scientific outreach event hosted by the British Academy in London in June 2019, featuring a live ultrasound demonstration (see [11] for a similar approach). This event was attended by many visitors local to the area, making the sample geographically homogeneous, and there was a relatively good balance of age and gender, which allowed us to explore the role of these factors in conditioning /r/-variation.

2. METHOD

We recorded the data while delivering a live ultrasound demonstration at a scientific outreach event. Visitors had the opportunity to try ultrasound tongue imaging (UTI), and they were invited to participate in the study. 55 visitors agreed to participate, 36 of whom were included in the current study. The inclusion criteria were having being born in and grown up in South of England. 18 participants described their gender as male (mean age = 38.4, SD =14.0), and 18 described it as female (mean age = 36.1, SD =17.7).

The participants read the following sentences: *That Mary Rivers! She's so overbearing. I'd feel better if I never saw her again. If you're not feeling well, a nice hot cup of tea can help you bear it. Try* *it out and see how you feel.*

The sentences were designed such that most (though not all) /r/ tokens occurred in a stable segmental environment, following a mid vowel and preceding a high vowel. The prosodic and morphological context was varied, including a word-initial onset (*Rivers*), a word-medial onset (*Mary*), morpheme-final /r/ (*overbearing*) and linking /r/ (*bear it*). We also aimed to elicit prosodically natural speech with these stimuli.

One of the experimenters held the ultrasound probe under the participant's chin along the midsagittal line. No head stabilisation was used. A lapel condenser microphone by AudioTechnica was used to capture the accompanying audio. Audio signal and the ultrasound image were recorded and synchronised using Articulate Assistant Advanced (AAA) v.2.18 [12]. The ultrasound system was EchoB, with a 2-4MHz probe. The frame rate was 60 frames per second, the depth was 75 mm and the number of scanlines was 127.

We first classified a selection of tongue images (at least one from each speaker) qualitatively, in order to identify the main types of rhotics in the data. We followed the classification system by [7]. The classification was done collectively by the three authors, who discussed the shapes and arrived at a consensus classification.

We then traced the tongue contour in the ultrasound image using Deep Lab Cut (DLC) [13]. This algorithm is trained on manually selected labels intended to pick out consistent flesh points in a midsagittal sequence, either with direct reference to articulatory features evidenced in the image (not limited to surface edge features), or through interpolation. The DLC process is trained therefore to assign the same labels to novel data. These tongue surface labelled points are located by DLC as part of a surface contour (a 2D spline) based on all the data in the image in reference to the training set provided by AAA (v.2.20). This surface contour can be exported as a series of Cartesian or polar points relative to an origin (e.g. the location of the short tendon).

Using a method based on [13], we measured the distance between the short tendon and the tongue surface along 11 points on the surface of the tongue. Approximate location of the points is illustrated in Figure 1. The radials connecting the points on the tongue surface and the short tendon form a fan, similar to a standard approach in ultrasound methodology. The distances between the short tendon and the tongue surface form a normalised tongue shape that can be compared across speakers.

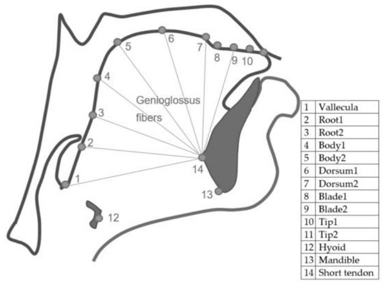

Figure 1: The reference points used in the analysis. Figure reproduced from [13] with permission.

We extracted the normalised tongue shapes for a single ultrasound frame in each /r/-token, at the time point when the tongue was judged to reach the maximum /r/-constriction. These representative UTI frames were selected by the first author, based on visual inspection of the ultrasound videos.

We fitted a series of GAMMs (Generalised Additive Mixed Models) modelling normalised tongue shapes, by predicting the distance from the tongue surface to the short tendon, depending on a by-fanline smooth. A stepwise model comparison procedure was used to evaluate the following predictors:

- main effect of gender;
- a smooth term for fanline by gender;
- a smooth term for participant age;
- tensor product interaction between the fanline and participant age;
- tensor product interaction between the fanline and participant age by gender;
- main effect of *r*-token;
- a smooth term for fanline by *r*-token.

The number of knots for each smooth was set to 10. All models included a tp by-participant random smooth for fanline with 10 basis functions [14]. Significance of the individual predictors was established through Maximum Likelihood comparisons between nested models. We selected the best model using a backward comparison procedure. We then corrected for autocorrelation by fitting an AR1 version of the selected model [15].

3. RESULTS

We found instances of the following /r/-types reported by [7]: front-up, tip-up and curled-up, as well as front-bunched /r/. We also found some tokens tentatively classified as mid-bunched, although a more appropriate label would perhaps be ‘weakly bunched’. Representative examples of all types are illustrated in Figure 2. The left column of the figure shows tongue outlines in Cartesian coordinates. The right column shows the corresponding normalised tongue shape, i.e. the distance from the short tendon to the tongue surface, depending on fanline. All these examples are from word-initial /r/ (*Rivers*). 13 out of 36 participants (36%) had tip-down /r/ in this position.

The distance measure used to produce normalised tongue shapes highlights that front-up, tip-up and mid-bunched /r/ are characterised by various degrees of convexity in the tongue surface, but they differ by the location of the constriction, which is more anterior for mid-bunched /r/. Additionally, tongue tip is down in mid-bunched /r/, which is reflected in the reduced distance at fanline 11. Curled-up /r/ and front-bunched /r/ both have a degree of tongue body concavity, which results from a simultaneous tongue root and dorsal constriction. The tip is notably lower in front-bunched /r/, compared to curled-up /r/.

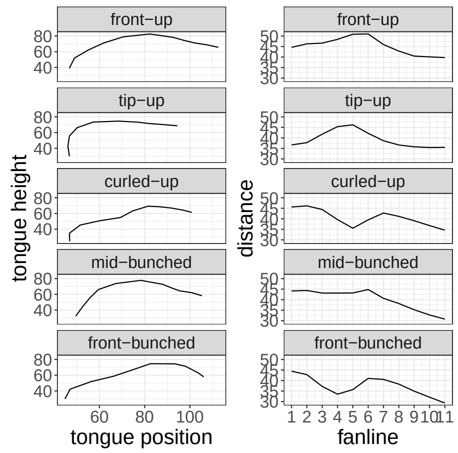

Figure 2: Left column: tracings of the tongue contour for word-initial /r/, as pronounced by three selected participants. Tongue tip is on the right. Right column: distance between the short tendon and the 11 reference points on the tongue surface for the same /r/-tokens.

Based on model comparison, normalised tongue shape was significantly affected by speaker age ($p<.05$) and by the /r/ token ($p<.001$), but not by speaker gender ($p=.07$). Gender had a significant main effect on distance ($p<.001$), whereby male speakers had overall greater distances compared to females, consistent with males having on average larger vocal tracts.

Figure 3 illustrates the model predictions for normalised tongue shape in word-initial /r/ (*Rivers*), depending on the fanline and speaker age. The figure shows predictions for female speakers (recall that males are predicted to have greater distance values but not significantly different shape). The mean smooths for various age groups separated by decade would appear to show a continuous progression from a more convex to a more concave tongue body, associated with a secondary dorsal constriction in speakers of ca. 50 years of age and younger. However, the only area of significant difference is fanline 11 (tongue tip), which suggests that variation in the tongue body is somewhat idiosyncratic. The tongue tip, on the other hand, shows significant lowering in apparent time, with youngest speakers being more likely to have a tip-down /r/.

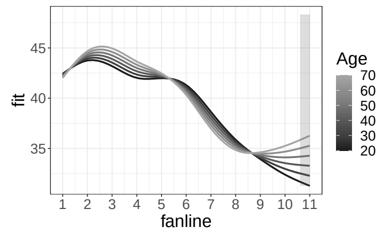

Figure 3: GAMM predictions for mean normalised tongue shape, depending on speaker age. Predictions are for a word-initial /r/ (*Rivers*) produced by a female speaker. Shading indicates areas of significant difference.

In terms of token effects, we find systematic differences in the tongue root and degree of dorsal constriction. Word-initial /r/ (*Rivers*) had more advanced tongue root compared to morpheme-medial /r/ (*Mary*), which in turn had more advanced tongue root and increased dorsal constriction, compared to morpheme-final one (*overbearing*) or linking /r/ (*bear it*). Morpheme final /r/ had a higher tongue blade, compared to linking /r/. While some systematic trends seem to emerge, they might be partially due to long-distance co-articulation. The rhotics in *her again* and *try* were markedly different, presumably due to the segmental context, and especially affrication in the latter. Otherwise, token-conditioned differences we find are relatively subtle. While some speakers showed forms of /r/-allophony, only few used clearly different /r/-shapes in different contexts. Typically, speakers had a single dominating /r/-shape, although we acknowledge that we tested a limited set of possible /r/ environments.

4. DISCUSSION

In some aspects, our results are consistent with previous studies on /r/-variation, crucially [7]. We find a mixture of various /r/-shapes, and an overall preference for tip-up /r/. A third of the speakers in our sample (36%) had a tip-down /r/ in their repertoire, compared to 42% in [7]. However, our results show an apparent time difference, suggesting an ongoing sound change from a tip-up to tip-down /r/ in the South of England. While, there is variation in tongue shape across all ages, younger speakers (under 25) are more likely to have a tip-down /r/ than a tip-up one.

From a certain point of view, this might seem surprising, as tip-up /r/ is generally the preferred variant in non-rhotic varieties of English [7, 11]. In general, tip-down /r/ is more likely in postvocalic positions [6, 10], so speakers who do not produce coda rhotics, are thought to be less likely to acquire bunching as a rhotic articulation [11]. We might therefore ask what could be driving a potential sound change towards tip-down /r/ in a non-rhotic variety.

We believe that a factor potentially contributing to this change is lip posture. Anglo-English /r/ has a prominent labial component, which has been shown to co-vary with tongue shape: tip-down /r/ is typically produced with more lip protrusion, compared to tip-up /r/, as shown by [7], who propose that lip protrusion is a compensatory strategy that allows speakers to lower the F3 despite the relatively small sublingual cavity produced by bunching the tongue (compared to retroflex). This explanation is consistent with their finding that various /r/ shapes are not associated with significant differences in the third formant, a key acoustic correlate of rhoticity. A further study has shown that Anglo-English listeners are highly sensitive to the visual cues associated with labialisation /r/ [16]. Visual information alone was sufficient for participants to disambiguate /r/ and /w/ minimal pairs, whereas an auditory classification task showed some confusability, with a general bias for /r/ perception. This was attributed to the presence of labio-dental [ʋ] as a frequent variant of /r/ in Anglo-English, which is an /r/-allophone with a relatively high F3.

Given that labialisation is a distinct feature of tip-down /r/, and also a prominent perceptual cue of rhoticity in Anglo-English, the labial gesture likely contributes to the propagation of tip-down /r/ in England [16]. A question that arises in this context is whether the shift towards tip-down /r/ we observed is a part of ongoing increase of labio-dental [ʋ]. This interpretation would be consistent with the presence of some weakly bunched /r/ tokens in our data, and also with the fact that we do not find a significant apparent time difference affecting tongue body. While we see an apparent time increase in tip-down /r/, we do not find an accompanying shift to a concave tongue body, which suggests that the emerging tip-down /r/ variants are not uniformly front-bunched. However, this proposal is only tentative, since we are not in a position to support it with acoustic or auditory evidence. We collected the data in a public space with significant background noise present, and as a result the audio data we have do not lend themselves to either spectral or auditory analysis.

Nevertheless, our results provide a compelling case for future studies of articulatory /r/-variation in South of England, also highlighting the need for stratifying the speaker sample for age. The current findings are among few that show systematic social structure in the variation between tip-up and tip-down /r/ in an accent of English. This type of variation is known to be conditioned by linguistic factors such as syllable position and segmental environment [6, 7, 11, 10], as well as by the variety of English itself, but otherwise, it can be idiosyncratic: some speakers have a bunched /r/ and some do not. The somewhat free nature of this variation is sometimes attributed to the acoustic similarity of retroflex and bunched /r/, especially at the frequencies typically salient for phonological contrasts [17]. A notable exception, however, is Scottish English, where /r/-variation is systematically conditioned by social class [8, 9].

Future articulatory studies may benefit from the approach we used to normalise tongue shape, originally proposed by [13]. A key aspect of the methodology is the use of a fan that uses an anatomically informed original point. Thanks to this, the distance measures can be interpreted in terms of articulatory constrictions, and they can be compared systematically across different speakers. Furthermore, by being fitted separately to each frame based on identifiable landmarks, the fan partially corrects for probe rotation and probe movement. While this approach is highly promising, it also raises questions concerning the level of systematicity one can expect when aggregating articulatory data. For example, we note some differences in the exact location of constriction for different types of /r/-variants (Figure 2). It is not clear whether such differences systematically distinguish /r/-variants, which would make them valuable for automatic classification of /r/-shape, or whether they are partially dictated by anatomical differences. Future models may improve on this approach by predicting tongue shape based on functional muscle bundle control.

5. ACKNOWLEDGMENTS

We thank our participants for contributing to our study. We also thank the British Academy for inviting us to participate in the Summer Showcase 2019, and for their support for our data collection during the exhibit.

6. REFERENCES

[1] P. Carley, I. M. Mees, and B. Collins, *English phonetics and pronunciation practice*. Abingdon, UK/New York, NY: Routledge., 2017.

[2] R. Ogden, *An introduction to English phonetics*. Edinburgh University Press, 2009.

[3] J. Wells, *Accents of English. 3 vols*. Cambridge: Cambridge University Press, 1982.

[4] P. Foulkes and G. J. Docherty, "Another chapter in the story of /r/: 'Labiodental' variants in British English," *Journal of Sociolinguistics*, vol. 4, pp. 30–59, 2000.

[5] E. Lawson, J. Stuart-Smith, J. M. Scobbie, and S. Nakai, "Dynamic Dialects: An articulatory web resource for the study of accents." [Online]. Available: https://www.dynamicdialects.ac.uk/

[6] P. Delattre and D. C. Freeman, "A dialect study of American *R*'s by X-ray motion picture," *Linguistics*, vol. 6, no. 44, pp. 29–68, 1968.

[7] H. King and E. Ferragne, "Loose lips and tongue tips: The central role of the /r/-typical labial gesture in Anglo-English," *Journal of Phonetics*, vol. 80, p. 100978, 2020.

[8] E. Lawson, J. M. Scobbie, and J. Stuart-Smith, "Bunched /r/ promotes vowel merger to schwar: An ultrasound tongue imaging study of Scottish sociophonetic variation," *Journal of Phonetics*, vol. 41, pp. 198–210, 2013.

[9] E. Lawson, J. Stuart-Smith, and J. M. Scobbie, "The role of gesture delay in coda /r/ weakening: An articulatory, auditory and acoustic study," *The Journal of the Acoustical Society of America*, vol. 143, pp. 1646–1657, 2018.

[10] J. Mielke, A. Baker, and D. Archangeli, "Individual-level contact limits phonological complexity: Evidence from bunched and retroflex /ɹ/," *Language*, pp. 101–140, 2016.

[11] M. Heyne, X. Wang, D. Derrick, K. Dorreen, and K. Watson, "The articulation of /ɹ/ in New Zealand English," *Journal of the International Phonetic Association*, pp. 1–23, 2018.

[12] Articulate Instruments Ltd. (2019) Articulate Assistant Advanced ultrasound module user manual, revision 2.18.

[13] A. Wrench and J. Balch-Tomes, "Beyond the edge: Markerless pose estimation of speech articulators from ultrasound and camera images using DeepLabCut," *Sensors*, vol. 22, p. 1133, 2022.

[14] M. Soskuthy, "Evaluating generalised additive mixed modelling strategies for dynamic speech analysis," *Journal of Phonetics*, vol. 84, p. 101017, 2021.

[15] R. H. Baayen, J. van Rij, C. de Cat, and S. Wood, "Autocorrelated errors in experimental data in the language sciences: Some solutions offered by Generalized Additive Mixed Models," in *Mixed-effects regression models in linguistics*, D. Speelman, K. Heylen, and D. Geeraerts, Eds. Springer, 2018, pp. 49–69.

[16] H. King and I. Chitoran, "Difficult to hear but easy to see: Audio-visual perception of the /r/-/w/ contrast in Anglo-English," *The Journal of the Acoustical Society of America*, vol. 152, pp. 368–379, 2022.

[17] X. Zhou, C. Y. Espy-Wilson, S. Boyce, M. Tiede, C. Holland, and A. Choe, "A magnetic resonance imaging-based articulatory and acoustic study of "retroflex" and "bunched" American English /r/," *The Journal of the Acoustical Society of America*, vol. 123, pp. 4466–4481, 2008.