The articulatory and acoustic impact of Scottish English /r/ on the preceding vowel-onset

Janine Lilienthal¹,²

¹ Speech and Hearing Sciences, Queen Margaret University, UK
² Centre for Speech Technology Research, University of Edinburgh, UK
jlilienthal@qmu.ac.uk

Abstract

This paper demonstrates the use of smoothing spline ANOVA and T tests to analyze whether the influence of syllable final consonants on the preceding vowel differs for articulation and acoustics. The onset of vowels either followed by phrase-final /r/ or by phrase-initial /r/ is compared for two Scottish English speakers. To measure articulatory differences of opposing vowel pairs, smoothing splines of midsagittal tongue shape recorded via ultrasound imaging are compared. For the acoustic data, differences of the first two formant frequencies at the onset are tested. The results confirm that there is no 1:1 mapping between articulation and acoustics.

Index Terms: anticipatory C-to-V coarticulation, ultrasound, smoothing spline ANOVA, Scottish English, rhoticity

1. Introduction

Scottish English is normally viewed as a rhotic dialect of English. Several studies, however, report changes to coda /r/, especially at phrase final position. For this position, Scottish English is increasingly transforming into a derhotic dialect. Derhotic variations were noticed for speakers from the Central Scottish Belt for a few decades [1, 2] and more recently [3, 4]. Derhoticization can be described as a phonetic weakening process and is linked with a dissociation of underlying /r/ gestures.

Acoustically, the sharp rise of the second formant frequency and a dip in the third formant, which are characteristic for rhotic /r/, are nearly invisible in the audio signal of derhotic speakers. Only traces of these characteristics may be found towards the end or after the voiced section. However, articulatory studies using ultrasound revealed that the underlying /r/ gestures, i.e., the raising of the tongue tip and the retraction of the tongue root, were not lost. Rather, analyzing articulatory and acoustic data simultaneously uncovered that the tongue tip gesture is delayed and appears right after the voiced section of the audio signal [5, 6, 7]. These findings of different gesture timing patterns for rhotic and derhotic speakers give rise to the question whether rhotic and derhotic realizations are different in their anticipatory coarticulation. Coarticulation is the overlapping of adjacent articulations and is caused by the transition of one articulation gesture to another involving different gesture timing patterns [8, 9].

To assess the impact of rhotic and derhotic realizations on coarticulation, we have performed an articulatory and acoustic analysis of Scottish English vowels that either precede a phrase final /r/ coda or that stand alone in coda position. We have used simultaneous ultrasound imaging and audio recording to measure the effect of anticipatory C-to-V coarticulation on the articulation and acoustics of rhotic and derhotic speakers. The analysis presented here is limited to the onset, as the onset is the most comparable time point for vowels vs. diphthongs. Ultrasound imaging enables us to capture midsagittal tongue shapes over time. It captures almost the whole tongue surface with some restrictions. For instance, a raised tongue tip would not be visible. However, in such case, information for the tongue blade may give sufficient information.

To determine whether tongue shapes for contrasting datasets have significant differences, we use smoothing spline analysis of variance (SS ANOVA), which allows us to visualize the differences between datasets by fitting a smoothed spline with a Bayesian confidence interval to each set of curves. Davidson [10] has previously described this method for ultrasound data. The advantage of SS ANOVA is that we not only can test whether there are significant differences in tongue shapes but, more importantly, locate the parts of the curves that differ.

For the acoustic data we compare the first two formant-frequencies at the onset of the vowel using the well-established T tests.

2. Methodology

2.1. Speakers and speech material

Ten female Scottish English speakers aged between 19 and 30 years were recorded. Two speakers, one from Glasgow (further as GLA, 23 years old) and one from Ayr (further as AYR, 20 years old), were chosen for the data analysis of the present study, as they best represent the variety of Scottish English /r/.

Speaker GLA was classified as rhotic, i.e., in the acoustic signal for /r#V/ and /rV#/ raising formant-frequency transitions of F2 were found in the second half of vowel-/r/ sequences. Speaker AYR on the other hand was classified as derhotic, i.e., diphthongization of vowel + /r/ sequences with traces of raising formant-frequency transitions for F2 were visible after the voiced section in the acoustic signal.

The speech material consisted of the six English monosyllabic words “beer”, “bee”, “bear”, “bay”, “bar” and “baa”. The words were integrated in carrier utterances like “This toothpaste is made of beer. Ebay sells 10 per minute.”. The target word was always phrase final and followed by a new phrase with either a word beginning with the same vowel /#V/ in case of syllable final /V/r/# or with initial /#rV/ in case of syllable final plain vowel /V/#. Twelve repetitions of each sentence in randomized order were read from a screen. Some tokens had to be excluded where speakers mispronounced or misread parts of the prompt. Thus, only ten tokens per vowel segment were analyzed.
2.2. Recording procedure and data analysis

During the recording, speakers wore a special stabilization headset to fix the ultrasound probe under the chin [11]. The microphone was attached to the headset. A midsagittal view of the tongue surface contour was recorded using a Mindray DP-6600 (Portable Ultrasound) with a frame rate of about 30 fps. For the audio signals, a sampling frequency of 22,050 Hz was used. Both, ultrasound images and audio signals were recorded at the same time using Articulate Assistant Advanced [12] (version 2.0.9), which allows for an approximate temporal synchronization of the two signals.

The vowel segments (/ar#/, /i#/ , /er#/ , /e#/ , /ar#/ and /a#/ ) of the utterance final words were labeled manually with Praat (version 5.0.46) and then imported to Articulate Assistant Advance and the EMU Speech Database (version 2.1.1). The vowel onset, i.e., the time point of articulatory and acoustic analysis, was placed at the onset of regular periodicity in the acoustic waveform. /ar#/ and /a#/ were excluded from the present report, as the formant frequency F2 in the first half of /ar#/ was considerably lower than for /a#/.

2.2.1. Articulatory analysis

Articulate Assistant Advanced was used to fit a spline to the tongue surface contour. The 42 knots controlling the spline are constrained so that each lies on one of 42 equally spaced radial axes. The knots were positioned by hand by tracing the tongue surface contour with the mouse. The tongue curves were captured as a series of x,y data points, scaled to actual size, in centimeter units with origin at the bottom left of the image. Each of the 42 spline-knots was exported. In addition, two interpolated points between each neighboring pair of knots were exported, giving 124 x,y points in total for each tongue contour.

To determine and quantify whether and at the same time where tongue shapes at the onset of /N/r# and /N/# show significant differences, a smoothing spline analysis of variance (SS ANOVA) was applied to the tongue data. The use of smoothing spline ANOVA to compare tongue curves of datasets with different conditions was recently introduced by Davidson [10]. Unlike Davidson, who uses the SS ANOVA implementation that comes with S-Plus 2000, we have used an equivalent implementation that is part of the ASSIST library [14] in R (version 2.8.0). An SS ANOVA entails two main stages. First, a smoothing spline, a type of natural cubic spline, that provides the best fit to all data points of a given set is computed separately for each dataset. Furthermore, a combined smoothing spline is computed for the combined data set that contains the points of the two data sets to be compared. The fitting is parameterized by a smoothing parameter which weighs the smoothness of the resulting spline versus the closest fit to all data points. The value of the smoothing parameter required to achieve a smooth spline is dependent on the noise level in the analyzed data set: the higher the noise, the bigger the required smoothing parameter. For the current analysis, generalized cross validation (GCV) was used to determine the smoothing factor for each data set. Additionally, a 95 % Bayesian confidence interval was computed for each smoothing spline to determine which sections of the curve show significant differences. Second, the interaction effect is computed for the datasets that are compared. The interaction effect for two datasets is the difference between the smoothing spline that represents the best fit of all data points of both datasets to the smoothing spline of each single dataset. Furthermore, 95 % Bayesian confidence intervals are computed to determine more precisely where significant differences are. A particular segment of the compared smoothing splines is classified as not significantly different from each other if the corresponding Bayesian confidence interval encloses zero on the y-axis.

2.2.2. Acoustic analysis

For the acoustic analysis, audio signals and acoustic annotations were read into the EMU Speech Database. Formant frequencies were calculated using the application tkassp in EMU. Formant errors were corrected manually. The formant frequencies F1 and F2 for the whole time course of vowel segments were exported to R and analyzed using the Emu-R library [13]. To compare the formant frequencies F1

![Figure 1: Results for the speaker GLA from Glasgow. The left column shows the result for /i/. The results for /e/ are given in the right column. The black plots represent /Vr#/ whereas the grey lines represent /V#. 95% Bayesian confidence intervals are visualized using dashed lines. Subfigure (a) gives the formant transitions for F1 and F2. The smoothing splines for the tongue shapes at vowel onset are given in Subfigure (b). Subfigures (c) and (d) show the interaction effects for /Vr#/ and /V#, respectively.](image-url)
and F2 of /Vr#/ and /V#/ show significant differences for /ir#/ and only little differences for /er#/.

For the tongue dorsum, we found different coarticulation effects for the rhotic and derhotic speaker. In case of the rhotic speaker, the tongue dorsum was clearly retracted for both /ir#/ and /er#/ compared to the contrasting vowels. For the Ayr speaker, the tongue root seems to be fronted. The tongue blade is raised for /ir#/ and /er#/ compared to the corresponding splines for /Vr#/, whereas /ir#/ (left) shows an undershooting compared to /i#, whereas /er#/ (right) shows an overshooting compared to /e#. Relative to the Glaswegian speaker, the compared smoothing splines for the Ayr speaker show even smaller difference (cf. Figure 2(b), (c) and (d)). For /ir#/ (left), no significant differences are visible at the tongue dorsum compared to /i#, whereas for /er#/ there are significant differences. It seems as if this part of the tongue moved slightly back, as well, even though the tongue root seems to be fronted. The tongue blade is raised for /ir#/ and /er#/ compared to the contrasted vowels.

An inter-speaker comparison shows that the gestures of tongue blade raising and tongue dorsum retraction are for both speakers visible at the vowel onset of /Vr#/ compared to its counterpart /V#. However, the rhotic speaker GLA shows greater differences in both parts than the derhotic speaker AYR.

4. Discussion

As the results show, rhotic and derhotic realizations of /i/ have no acoustic impact on the coarticulation of preceding vowels. For both speakers, no significant difference between the formant frequencies at the vowel onset can be found. This finding extends to approximately the first third of the vowel. For the articulatory data, the picture is different. The results show for both vowels /i/ and /e/ significant differences in the tongue curves. For both speakers, the tongue blade is raised at the onset of /Vr#/ compared to the onset of /V#. However, this effect is more pronounced for the rhotic speaker. For the tongue dorsum, we found different coarticulation effects for the rhotic and derhotic speaker. In case of the rhotic speaker, the tongue dorsum was clearly retracted for both /ir#/ and /er#/ whereas the derhotic speaker shows no significant differences for /ir#/ and only little differences for /er#/. The less pronounced differences between /Vr#/ and /V#/ articulation of the derhotic speaker might be a result of the difference of gesture timing for rhotic and derhotic speakers. Derhotic speakers show dissociated gesture patterns. However,
to confirm this hypothesis, a further study with more subjects is required.

Furthermore, we found that smoothing spline ANOVA is an adequate method to show coarticular effects in the articulation of vowel-onsets for differing conditions using ultrasound tongue imaging. SS ANOVA allows us to quantify differences of tongue shapes and identify differing segments of tongue curves.

For further experiments, it should be considered to vary the speech task. Our current experiment, i.e., reading randomized sentences from the screen, made most speakers not realize derhotic /r/s in the reading task. We attribute this effect to the reading of long utterances from a screen. Potentially, using a prompted task would provoke more spontaneous and casual speech and show higher rates of derhoticization.

5. Conclusion

We have presented first results for a pilot study on anticipatory coarticulation of coda /r/ on the preceding vowel. Two speakers were chosen out of a larger database, as they presented varieties of phrase final /r/ coda in Scottish English most. One speaker was classified as rhotic, whereas the other speaker was classified as derhotic.

For the acoustics analysis, we tested the first two formant frequencies of the vowel-onset of /N/r# vs. /N#. vs /N/# on significant differences. All compared vowel pairs for both speakers were not significantly different.

To investigate whether the articulation of contrasting vowel pairs differs at the onset, we have used smoothing spline ANOVA, which was applied to the tongue-curve data recorded via ultrasound. This method enabled us not only to show whether tongue shapes of contrasting conditions differ, but also where they differ.

Contrary to the findings of the acoustics, the articulatory analysis revealed that syllable final /r/ does have an impact on the onset of the preceding vowel. Both speakers showed significant differences of the tongue shapes at the onset of /N/r# sequences compared to the onset of /N#. A raised tongue blade was found for all vowels and speakers in vowel-onsets with following /r/ coda. For the tongue dorsum, however, we have found different results for the two speakers.

The tongue dorsum was retracted for the rhotic speaker, but not for the derhotic speaker. The derhotic speaker showed no significant differences for /ir# and only small differences for /er#.

Looking at the overall differences of the tongue curves, the rhotic speaker showed considerably greater differences between the contrasting tongue shapes than the derhotic speaker. This might be a result of different gesture timing, as derhotic speakers show dissociated gesture patterns. To confirm this hypothesis, a further study with an extended database of rhotic and derhotic speakers is needed. Also, it might be useful to employ a different speech task that, other than the currently used reading task, provokes more spontaneous speech.

Lastly, this report is a further proof that there is no 1:1 articulatory-to-acoustic mapping. Changes in the articulation need not necessarily be detected in the acoustic signal.

6. Acknowledgements

I would like to thank James Scobie, Sonja Schaeffler and Alan Wrench for their helpful comments and discussion. This research was funded by the Marie Curie EdSST program (MEST-CT-2005-020568).

7. References