Shape Coarticulation in the Spatial Frequency Domain: An Example Using /ɹ/

Katherine M. Dawson[1], Micalle Carl[1], D. H. Whalen[1,2,3]

[1]Graduate Center of the City University of New York, [2]Haskins Laboratories, [3]Yale University
kdawson2@gradcenter.cuny.edu, mcarl@gradcenter.cuny.edu, dwhalen@gc.cuny.edu

ABSTRACT

One facet of coarticulation is tongue shape assimilation to surrounding segments. This study demonstrates the use of the Fourier transform for quantifying the effect of rhotics on surrounding sounds. Ultrasound images taken from American English speakers producing a rhotic-heavy sentence are used as test data. The method shows promise for describing shape coarticulation without the need for head stabilization. Potential uses of this method involve quantification of speech motor control strategies among different groups of speakers, such as children acquiring language and speakers with disorders of the speech motor system.

Keywords: ultrasound; tongue shape; coarticulation; Fourier analysis

1. INTRODUCTION

1.1. Coarticulation

In speech production, coarticulation can be defined as the temporal and spatial overlap of speech gestures [5], such that a single speech segment may be realized differently, depending on the context of the surrounding gestures [11]. Various theories account for coarticulation, including those that focus on feature spreading and those that focus on coproduction (for a review, see [4]). Studying coarticulation provides insight into the nature of a speaker's speech motor control [11]. The coordination between coarticulating gestures may be impaired in individuals with speech disorders, impacting upon intelligibility [8, 19].

The study of coarticulation has been conducted in a number of modalities, including the perceptual (e.g. [20]), acoustic (e.g. [14]), and articulatory (e.g. [1]) domains. Acoustic quantification of coarticulation has been conducted within a variety of vowel and consonant contexts, including calculation of degree of articulatory constraint (DAC; [16]), and the locus equation [10, 18].

The connection between acoustics and articulation using the locus equation has been made by Iskarous, Fowler, and Whalen [9]. Articulatory measurement of coarticulation has employed techniques such as electropalatography, (EPG; e.g. [15]), electromagnetic articulography (EMA; e.g. [14]), and ultrasound (e.g. [12, 22]).

The liquid approximant /ɹ/ has been shown to exhibit considerable coarticulatory influence over surrounding consonants and vowels, over both a local and long domain [7], where coarticulated segments are described as having a rhotic-like quality. The rhotic also has a high degree of coarticulatory resistance to surrounding segments [2]. This is consistent with the DAC model, in which segments that have increased tongue dorsum constraint exhibit higher resistance to coarticulatory influences from surrounding segments [16], as well as greater coarticulatory aggression on adjacent sounds [14].

Determining the extent and nature of coarticulation requires direct measures of the articulators. Acoustic measures, such as lowered F3, can indicate rhoticity, but represent the contribution of more than one articulator. Articulatory techniques such as EMA and EPG provide insight by allowing measurement of articulator movements with high temporal resolution, but do not capture the shape of the tongue contour. Ultrasound imaging holds promise of showing more of the contribution of the tongue to /ɹ/ coarticulation by allowing visualization of a large section of the mid-sagittal or coronal tongue contour. Indeed, ultrasound imaging has been applied to the study of coarticulation in both children and adults (e.g. [23]), using analyses such as the mean nearest neighbour distance between tongue shapes for adjacent segments [24].

1.2. The current study

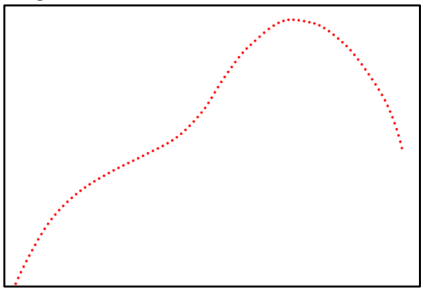
This study used ultrasound images of the tongue to quantify the degree of coarticulation exerted by American English /ɹ/. Tongue shapes for rhotics and nearby segments were transformed into the spatial frequency domain using a Fourier transform-based analysis adapted from Liljencrants [13]. This method seemed promising in terms of differentiating types of shape using their spatial frequency profiles. We hypothesized that /ɹ/ would influence segments both adjacent to it and at a greater remove, and that the Fourier transform would reflect this.

2. METHOD

2.1. Participants, stimuli and procedure

Four American English speakers were imaged using ultrasound while producing repetitions of the sentence ‘Let Robby cross Church Street’. This sentence contains /ɹ/ in four different contexts: onset position (/ɹɑbi/), CC (/kɹɑs/), syllabic (/ʧɝʧ/) and CCC (/stɹit/). Four repetitions of the sentence were analysed per speaker. The tongue contour was extracted from the image containing the maximal constriction for each tongue shape, as a sequence of one hundred *x,y* coordinate points, using a MATLAB-based script (see figure 1).

Figure 1: Contour points extracted from ultrasound image representing the mid-sagittal tongue surface.



Shape data from the four speakers were analysed to test the viability of the Fourier transform method for quantifying tongue shape coarticulation for all /ɹ/ shapes, plus the vowels in ‘Robby’, ‘cross’, ‘street’, and the /s/ segments in ‘cross’ and ‘street’. For all speakers we additionally analysed the vowels /ɑ/ and /i/ in a /bVb/ context and the consonant /s/ in a /ɑCɑ/ context, to compare to their counterparts in the sentence context. For one speaker (01_FC) we also analysed /ʌ/ in the context /bVs/, as this speaker produced a different vowel for ‘cross’ compared to ‘Robby’ and ‘bob’.

2.2. Fourier transform

The tongue shapes were analysed using the Fourier transform method adapted from Liljencrants [13] as follows:

- The tangent angle at each point on the tongue contour (see figure 1) was calculated, using a central difference method. This produced a tangent angle function for each shape.
- The Fast Fourier Transform (FFT) [3] was used to transform the tangent angle function into the spatial frequency domain.
- The 0th (d.c.) coefficient was discarded from this analysis, as it represents a constant offset of the function. The remaining coefficients are invariant to rotation of the original shape, as the effect of the rotation is contained in the 0^{th} coefficient.

A key difference between our analysis and that of Liljencrants [13] is that his study used a transform of positional data of the tongue contour rather than the tangent angle function. This was possible because his data were taken from X-ray images of the vocal tract, and could therefore be placed in a coordinate system with the origin and axes fixed with reference to the hard structures visible in the images. Our data had no such static reference points, so the transform of the tangent angle function was chosen as it gives a result which is invariant to scale and rotation. The result of this, however, is that the findings of Liljencrants (that phase and magnitude within the Fourier coefficient relate to constriction location and constriction degree, respectively) cannot be assumed to apply in this analysis.

3. RESULTS

3.1. Fourier transform

Figures 2 and 3 show the Fourier transform first coefficient (C1) values for each tongue shape, plotted on real and imaginary axes, for two of the four participants: 01_FC (female speaker) and 04_MC (male speaker). Each symbol is representative of a single instance of a tongue shape. The tongue shapes in the non-sentential context are surrounded by solid-line one-standard deviation confidence ellipses (e.g. /ɹ/ in a /ɑCɑ/ context). The context and segment type are given by the legend. The tongue shapes in the sentential context are in dashed one-standard deviation confidence ellipses.

Both figures demonstrate that all /ɹ/ shapes have a high value in the real part of C1. The /ɹ/ shapes in all contexts also overlap considerably for 01_FC. In contrast, the /ɹ/ shapes for 04_MC show less overlap. The onset /ɹ/ in ‘Robby’ and the /ɹ/ in /ɑCɑ/ context are closer for this speaker, whereas the other three contexts (‘cross’, ‘Church’ and ‘street’) are more similar. From inspection of figure 4, which shows /ɹ/ tongue shapes for all participants, it is clear that 04_MC has the highest level of variation in shape, and that the Fourier plot is reflecting this.

The /i/ tongue shape in /bVb/ context has a high value on the imaginary axis of C1. The /i/ shape in ‘Robby’ is lower on this axis, and the /i/ in 'street' is lower again (i.e. moving towards the /ɹ/ values). This same pattern is seen for the other vowels (/ɑ/ for 04_MC and /ɑ/ and /ʌ/ for 01_FC) and also /s/, in that they are shifted higher on the real axis. The patterns shown for these participants are consistent with the results of the two other participants, whose plots are not shown due to space considerations.

Figure 2: Fourier transform first coefficient (C1) plot for 01_FC

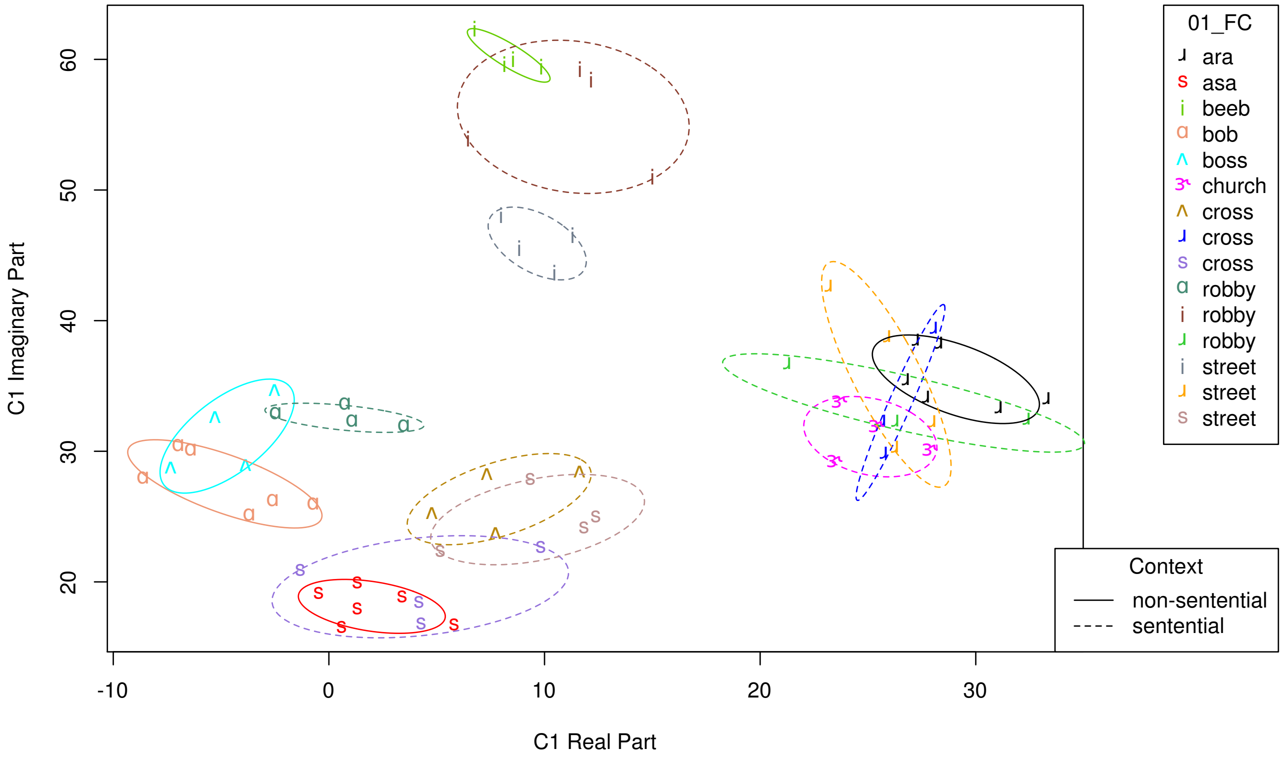


Figure 3: Fourier transform first coefficient (C1) plot for 04_MC

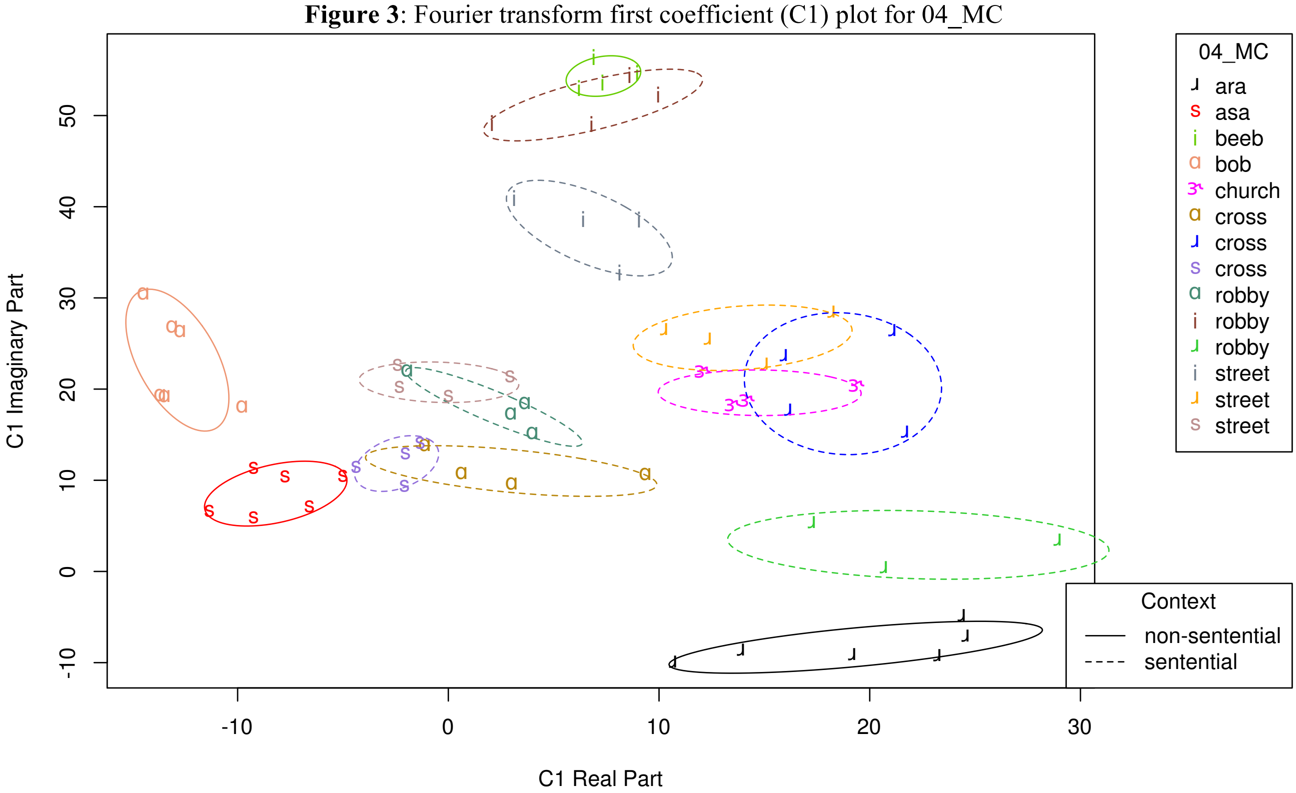
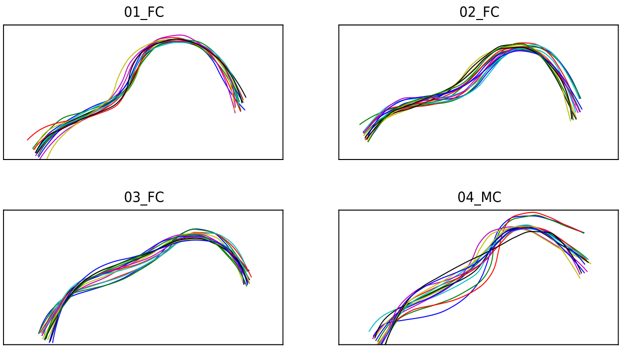


Figure 4: /ɹ/ shapes across all contexts for the four speakers.



4. DISCUSSION

The patterns observed in the graphs of the first coefficient of the Fourier transforms (Figures 2 and 3) indicate that this method tracks tongue-shape coarticulation. All non-rhotic tongue shapes in rhotic context have values closer to those of the /ɹ/ tongue shapes than they do in non-rhotic environments. Further, the contexts in which more coarticulation would be expected (i.e. segments adjacent to the rhotic) are shifted a greater amount towards the cloud of /ɹ/ shapes than the contexts in which less coarticulation would be expected (e.g., for /s/: 'street' > 'cross', for /i/: 'street' > 'Robby').

For 01_FC, the /ɹ/ shapes are altered relatively little in variable contexts. For 04_MC, this is not the case. The two other participants (not illustrated) fall between these two extremes. More information on shape coarticulation for a greater number of participants than used in this study may be useful in adding extra dimensions to theories of coarticulation, such as the DAC model [16].

The magnitude of the coefficient (distance of the segment types from the origin) appears to correspond roughly to the extent of the inflection of the tongue body, in that /i/ has the highest values for all contexts across all participants. This seems plausible in that C1 is the coefficient of the lowest spatial frequency of the shape, and as such, relates to the largest-scale features. We would also expect this to be true for segments such as /g/ and /u/, which have a single, extreme tongue body inflection. Values for segments such as /s/ show C1 coefficients with a low magnitude, as, despite the high degree of constriction, the shape is not highly inflected.

In general, we can say that the coarticulatory influences on /i/ in this data set appear to be acting in the magnitude dimension (i.e. reduction of the magnitude of the coefficient, most likely relating to reduction of the inflection). However the coarticulatory influences on /ɑ/, /s/ and /ɹ/ are acting in both phase (the angle from the *x*-axis to a line connecting the segment to the origin) and magnitude. As a result, the level of coarticulation cannot be quantified in a single parameter. If an *a priori* decision was made as to the dimension of interest (e.g. magnitude for /i/), then movement could be quantified in this dimension. Further explorations of quantification of the amount of coarticulation would require larger data sets; these are in the process of being collected. The relationship of the shape changes to the amount of acoustic evidence of coarticulation is also under investigation.

A caveat to this analysis is that the sentence we used - 'let Robby cross Church Street', contains four instances of the rhotic. This makes it difficult to attribute coarticulatory effects to one of the rhotics exclusively. The assumption was made that the nearest rhotic would exert the most coarticulatory influence, and comparisons were made with rhotics in a non-sentential context. Ideally, tongue shapes in the context of each of these rhotics would be studied in a number of sentences where each rhotic was the only one present. However, we believe that the current study demonstrates the viability of this method for studying tongue shape coarticulation, and further research can focus on quantifying the coarticulatory effects of rhotics in different contexts, on different segments and among different populations.

Adapting this method to purely shape-based data, where no meaningful scale or orientation of the tongue shape is available makes it applicable in a wider range of experimental settings than those that require head stabilization [17] or compensation for movements of the ultrasound probe and head [21]. The technique may be particularly useful when studying populations for whom head correction or restraint is difficult or undesirable, such as children and those with dysarthria. More generally, the Fourier analysis provides an easily calculated metric for an assessment of coarticulation as evidenced in tongue shape.

5. ACKNOWLEDGMENTS

The authors gratefully acknowledge support from NIH Grant DC-002717 to Haskins Laboratories and the City University of New York.

6. REFERENCES

[1] Bell-Berti, F., Harris, K. S. 1981. A temporal model of speech production. *Phonetica,* 38, 9-20.

[2] Boyce, S., Espy-Wilson, C. Y. 1997. Coarticulatory stability in American English /r/. *J. Acoust. Soc. Am*. 101(6), 3741–3753.

[3] Cooley, J. W., Tukey, J. W. 1965. An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation* 19(90), 297-301.

[4] Farnetani, E., Recasens, D. 1999. Coarticulation models in recent speech production theories. In W. J. Hardcastle & N. Hewlett (eds), *Coarticulation: Theory, Data, and Techniques*. New York: Cambridge University Press, 31-65.

[5] Fowler, C. A., Saltzman, E. 1993. Coordination and coarticulation in speech production. *Language and Speech* 36, 171–195

[6] Gick, B., Bernhardt, B., Bacsfalvi, P., Wilson, I. 2008. Ultrasound imaging applications in second language acquisition. In J. Edwards & M. Zampini (eds), *Phonology and second language acquisition*. Amsterdam: John Benjamins, 309-322.

[7] Heid, S., Hawkins, S. 2000. An acoustical study of long domain /r/ and /l/ coarticulation. *Proc. 5th Seminar on Speech Production: Models and Data.* Kloster Seeon, Germany, 77-80.

[8] Hodson, S. L., Jardine, B. R. 2009. Revisiting Jarrod: Applications of gestural phonology theory to the assessment and treatment of speech sound disorder. *International Journal of Speech-Language Pathology* 11(2), 122–134.

[9] Iskarous, K., Fowler, C. A., Whalen, D. H. 2010. Locus equations are an acoustic expression of articulator synergy. *J. Acoust. Soc. Am.* 128, 2021-2032.

[10] Krull, D. 1987. Second formant locus patterns as a measure of consonant-vowel coarticulation. *Phonetic Experimental Research at the Institute of Linguistics, University of Stockholm, Sweden*, *V*, 43–61.

[11] Kühnert, B., Nolan, F. 1999. The origin of coarticulation. In W. Hardcastle & N. Hewlett (eds), *Coarticulation: Theory, Data, and Techniques*. New York: Cambridge University Press, 7-30.

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

[13] Liljencrants, J. 1971. A Fourier series description of the tongue profile. *Speech Transmission Laboratory Quarterly Progress and Status Report*, *12*(4), 9–18.

[14] Recasens, D., Espinosa, A. 2009. An articulatory investigation of lingual coarticulatory resistance and aggressiveness for consonants and vowels in Catalan. *J. Acoust. Soc. Am*. 125(4), 2288–98.

[15] Recasens, D., Pallares, M. D. 1999. A study of /r/ and /r/ in the light of the “DAC” coarticulation model. *J. of Phonetics* 27, 143–169.

[16] Recasens, D., Pallares, M. D., Fontdevila, J. 1997. A model of lingual coarticulation based on articulatory constraints. *J. Acoust. Soc. Am*. 102(1), 544–561.

[17] Scobbie, J., Wrench, A., van der Linden, M. 2008. Head-probe stabilisation in ultrasound tongue imaging using a headset to permit natural head movement. *Proc 8th ISSP,* Strasbourg, France, 373-376.

[18] Sussman, H. M., Bessell, N., Dalston, E., Majors, T. 1997. An investigation of stop place of articulation as a function of syllable position : A locus equation perspective. *J. Acoust. Soc. Am*. 101, 2826–2838.

[19] Tjaden, K. 2000. An acoustic study of coarticulation in dysarthric speakers with Parkinson’s disease. *JSLHR* 43, 1466–1480.

[20] West, P. 1999. Perception of distributed coarticulatory properties of English /l/ and /r/. *J. of Phonetics*, 27(4), 405–426.

[21] Whalen, D. H., Iskarous, K., Tiede, M. T., Ostry, D., Lehnert-LeHoullier, H., Hailey, D. 2005. HOCUS, the Haskins optically-corrected ultrasound system. *JSLHR* 48, 543–553.

[22] Zharkova, N., Hewlett, N. 2009. Measuring lingual coarticulation from midsagittal tongue contours: Description and example calculations using English /t/ and /a/. *J. of Phonetics* 37(2), 248–256.

[23] Zharkova, N., Hewlett, N., Hardcastle, W. J. 2008. An ultrasound study of lingual coarticulation in children and adults. *Proc 8th ISSP,* Strasbourg, France, 161–164.

[24] Zharkova, N., Hewlett, N., Hardcastle, W. J. 2011. Coarticulation as an indicator of speech motor control development in children : An ultrasound study. *Motor Control* 15, 118–140.