Coarticulatory Stability in American English /r/

Suzanne Boyce and Carol Y. Espy-Wilson

ECS Engineering Department, Boston University, Boston, MA 02215

ABSTRACT

A number of different researchers have reported a substantial degree of variability in how American English /r/ coarticulates with neighboring segments. Acoustic and articulatory data were used to investigate this variability for speakers of "rhotic" American English dialects. The major issue addressed is the degree to which segmental context affects articulatory movement as reflected in the F3 trajectory. In particular, we ask whether the duration of the F3 trajectory is affected by conflicting vs. nonconflicting articulatory specifications. The F3 formant trajectory durations were measured by automatic procedure and compared for nonsense words of the form /'waCrav/ and /waC'rav/, where C indicates a labial, alveolar or velar consonant. These durations were compared to F3 trajectory durations in /'warav/ and /wa'rav/. Results indicated F3 trajectory durations were similar across consonant contexts, suggesting that coarticulation of /r/ is achieved by overlap of a stable /r/-related articulatory gesture with gestures for neighboring sounds. This interpretation, and the concordance of F3 time course with tongue movement for /r/, was supported by direct measures of tongue movement for one subject.

1. Introduction

This paper is concerned with acoustic and articulatory aspects of the way consonantal /r/ (as classically defined) interacts with adjoining consonant and vowel segments in “rhotic” varieties of American English. Because /r/ as produced by American English speakers shows wide variability in articulatory configuration, we concentrate on analysis of consistency in its acoustic signature. We are concerned with two types of coarticulatory situations. In the first, we consider anticipatory coarticulation of /r/ coloring into adjacent segments with non-conflicting specification for tongue placement. In the second, we consider anticipatory coarticulation of /r/ coloring into adjacent segments whose articulatory specifications might conflict with the articulatory specifications for /r/. In addition, we consider the effect of stress.

The most salient feature of American English /r/, whether consonantal or vocalic, is its low F3, which can range between 1100 and 2000 Hz but which is normally in the region of 1600 Hz for both men and women [1][2][3]. For other segments of American English, F3 ranges approximately between 2100 and 3000 Hz [4]. In general, it is reasonable to assume that the time course of frequency change in F3 below 2000 Hz reflects the time course of articulatory movement specific to /r/.

1.1. Models of Coarticulation

Classically, coarticulation is defined as an assimilation in the articulation of one segment–a “target” segment–as a result of a neighboring “home” segment. Physically, coarticulation may be manifested as a change in dynamic characteristics of movement (shape/displacement/duration of the articulatory movement) as well as change in placement within the vocal tract. Treatments of coarticualtion in the literature have typically contrasted models that predict a) increased duration of articulatory movements in the presence of neutral or nonconflictings segments [5][6][7] and b) stability in the articualtory movements when nonconvlicting segments intervene ([8][9][10][11] among others).

2. Methodology

2.1. Stimuli and speakers

Seven speakers (three female and four male) produced 5 repetitions of experimental nonsense words /wɑvrɑv/, /wɑbrɑv/, /wɑgrɑv/, /wɑdrɑv/, and 5 repetitions of the control nonsense words /wɑrɑv/, /wɑwɑv/, /wɑvɑv/, /wɑbɑv/, /wɑg ah v/ and /wɑdɑv/. Each nonsense word was produced in two stress conditions: with stress on the first syllable (initial stress) and with stress on the second syllable (final stress); e.g. /wɑ'rɑv/, /'wɑrɑv/. All words were embedded in the carrier phrase “Say ____________ for me”. For 6 of these 7 speakers, productions were recorded in a quiet room using a pressure-gradient close-talking noise-cancelling microphone (part of Sennheiser HMD 224X microphone/headphone combination). The subjects produced the experimental stimuli in the same order 5 times with reference to a handheld paper list. The utterances were digitized at 16 kHz on a SUN workstation.

Articulatory plus acoustic data were obtained from a seventh speaker, RD, who also produced the same nonsense words with the exception of /wɑbɑv/ and /wɑbrɑv/. For this speaker, movement of 2 electromagnetic transducers placed on the tongue tip and tongue dorsum was recorded via an Electro-Magnetic Midsagittal Articulometer (EMMA) apparatus as in [12]

Subjects were speakers of fully rhotic versions of standard American English from Missouri, Western Massachusetts, Upper New York State, Western Pennsylvania, Michigan, Philadelphia and Washington State. Speakers were instructed to produce words at a self-selected comfortable and consistent rate, in a natural manner, and were given a short practice session.

The experimental nonsense words were designed to include cases with labial, alveolar, and velar consonants before /r/. The control words were included to allow analysis of the formant trajectories characteristic of these consonants as well as those of the labial most like /r/ (/w/) and of /r/ itself. Additionally, the comparison between /g/ and /w/ provided a rough indication of the extent of F3 lowering attributable to rounding.

The experimental words /wɑbrɑv/ and /wɑvrɑv/ were expected to present minimal barriers to coarticulation of /r/; that is, we expected that if the duration of ariculatory movement increases in nonconflicting environments, these words would show longer F3 trajectories (and presumably longer articulatory trajectories) than those for words with singleton /r/. If articulatory movements for a segment are spatiotemporally stable, as predicted by the coproduction model, then we expected trajectories to be the same as those for words with singleton /r/. The words /wɑgrɑv/ and /wɑdrɑv/ were expected to present articulatory barriers to coarticulation because they involve the tongue.

F3 trajectories were used throughout the study as an index of articulatory movement associated with /r/. Because portions of these acoustic trajectories were sometimes hidden by acoustic effects such as aspiration, lack of glottal vibration, etc., the point of trajectory minimum was not always represented measurably in the signal. Thus, the object of measurement was always the full trajectory including F3 lowering, extremum and rising components.

Formant tracks were computed for all the utterances using the ESPS/WAVES formant tracker and a 10 ms frame rate. Alignment between formant tracks and spectrograms was handled automatically as part of the WAVES program. In cases where acoustic and articulatory data were compared directly, data acquisition frame rates were matched by recomputing the formant tracks with a 51.2 ms window and a 3.2 ms frame rate. These were aligned by taking into account a shift of half the window length. The formant tracks were edited by the two authors working together to eliminate noisy or erroneous data points as described below. At least 3 tokens were analyzed for each speaker and each word.

2.2. Editing Results from Formant Tracker

A number of issues came up in the course of editing the results of the formant tracker. For instance, sometimes stops were produced with incomplete vocal tract closure, and the formant tracker was able to detect consistent and appropriate F3 values in at least some portion of the acoustically defined closure interval. In some cases, the formant tracker incorrectly assigned values belonging to F3 as belonging to F2 or F4; these values were replaced by the correct values. Finally, there were cases in which the formant tracker identified energy simultaneously at two points in the spectrum which might plausibly reflect F3. Almost invariably in these cases of "double" paths, the upper track resembled the pattern of F3 seen during the closure in control words /wɑdɑv/, /wɑgɑv/, /wɑbɑv/, /wɑvɑv/, while the visible portion of the lower track resembled the pattern seen for /r/ in /wɑrɑv/. Because the articulatory data paralleled the lower track, we took this path as reflective of the F3 time course specific to/r/. (See Figure 2b for an example.)

All editing was done by visual reference to spectrograms for each token with results of the formant tracker superimposed, and, where appropriate, power spectra at selected points during consonant closure. Several steps were involved in editing the formant tracks. First, F3 tracks during the word-initial /w/ and word-final /v/ were deleted. If the F3 tracks during the intervocalic obstruents were noisy, the frequency values were deleted while maintaining the correct spacing in time.

2.3. Measure of /r/ trajectory duration

An automatic procedure was developed to measure the duration of the /r/-related F3 trajectories. We defined F3 trajectory duration as the time lapse between the inflection point at trajectory beginning (on the left side) and inflection point at trajectory end (on the right side). Our program found these inflection points based on the values of the first and second differences of the F3 trajectory, first on the left side of the F3 minimum and then on the right side of the F3 minimum (see Figure 1). Pertubations were smoothed by hand and missing values eliminated during editing were filled in by a simple linear interpolation algorithm, producing a continuous trajectory. We estimate error at +/- 20 ms.

3. Results

Although we focused on duration across context, interestingly /r/-related F3 trajectories across our data showed striking similarity in shape as shown in Figure 1.

The F3 duration values were entered into analyses of variance

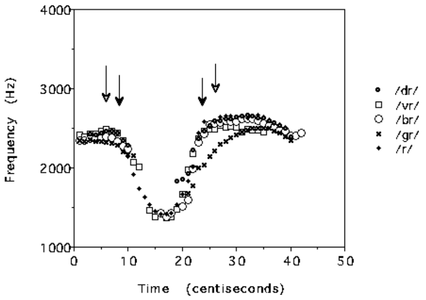

Figure 1: F3 tracks taken from one token each of /wɑ'rɑv/, /wɑv'rɑv/, /wɑb'rɑv/, /wɑg'rɑv/ and /wɑd'rɑv/. Inflection points for F3 tracks of /wɑ'rɑv/ and /wɑv'rɑv/ are marked with arrows.

using the factors SPEAKER and CONTEXT (/b/, /d/, /g/, /v/, or /r/). Since our database included words with both initial and final stress, STRESS was also a factor. The major hypothesis considered was whether F3 trajectory duration differs across contexts. Because of correlations naturally existing across data from particular subjects, particular items, and particular stress patterns, we elected to treat each of these factors as a correlated variable in a repeated measures analysis of variance. Separate "subject" and "item" repeated measures analyses of variance were performed using, respectively, subject variability, consonant context variability and stress variability as the error term. In the subject analysis, context and stress were used as "within", or "repeated" measures and SPEAKER was a "between" or "grouping" factor. Two item analyses were performed. In one, CONTEXT was used as a "between" factor and SPEAKER as a "within" factor. In the second, STRESS was used as a "between" factor, and SPEAKER as a "within" factor. Because for subject RD data from /br/ was not collected, the items analysis CONTEXT factor included /vr/, /dr/, /gr/, and /r/. The dependent variable in all cases was DURATION of the F3 trajectory. Individual cells were represented in the subject analysis by tokens; in the item analyses cells consisted of token means for a particular context or stress condition.

Results showed CONTEXT was not significant in either analysis (Subject: df = 4, F = 2.18, p> .10; Item: df = 3, F = .140, p>.50), suggesting that F3 trajectory duration was consistent regardless of whether /r/ was the single intervocalic consonant, or whether it followed /b/, /v/, /d/ or /g/. SPEAKER (Item: df = 6, F =19.9, p<.0001) and STRESS (Subject: df =1, F= 5.2, p<.05; Item: df =1, F = 43.5, p<.001) were significant. Interactions (STRESS x CONTEXT, SPEAKER x CONTEXT and STRESS x SPEAKER) were not significant suggesting that the result of stable trajectory durations across context was consistent across stress and speaker.

The SPEAKER effect reflected a tendency for different speakers to show characteristically longer or shorter F3 trajectories. Stress was significant but this appeared to be due primarily to measurement anomalies with the automatic procedure (see [13]).

4. Conclusion

The data in this study demonstrate that F3 trajectories for /r/, for any one subject, show relatively consistent duration and shape across a number of variables that might be expected to affect the way /r/ is articulated. Notably, the similarity in trajectory shape indicates that duration for all components of F3 trajectories–onset (lowering), extremum and offset (raising)–remains consistent across phonetic context. Similarly, duration of the full F3 trajectory is consistent across phonetic contexts. Thus, when overlapping influences are eliminated, it appears that whether the segment preceding /r/ is alveolar, velar, labial or vocalic does not affect the essential shape or duration of the F3 trajectory. These results support models of coarticulation (e.g. the coproduction model) that emphasize stability in articulatory movements across contexts.

It is interesting to speculate on the articulatory mechanisms of trajectory duration maintenance for /r/. This may be related to the well known fact that American English speakers show a number of different articulatory configurations for /r/ roughly categorized as "retroflex" and "bunched" [14]. One possibility for trajectory maintenance is that subjects "swap" between bunched and retroflex articualtion s of /r/ according to the requiremnts of context. Alternatively, it is possible that constriction location is less important than some other manipulation of the vocal tract affecting resonance. More research is needed to elucidate this question.

5. ACKNOWLEDGMENTS

The articulatory data discussed in this paper were collected with the generous help of Dr. Joseph Perkell, Dr. Melanie Matthies, Dr. Mario Svirsky and Jane Wozniak. We are indebted to them and to many others for comments and suggestions, including Dr. Kenneth Stevens, Nabil Bitar, Demetrios Paneras, Dr. Caroline Huang, Dr. Sharon Manuel and Dr. Rena Krakow. This work was supported in part by: NSF Grant #9296032, NSF Grant #IRI-9310518, a Clare Boothe Luce fellowship to the second author, and a Boston University College of Engineering Dean's postdoctoral fellowship and NIH Grant #1R030C-02576 to the first author.

6. REFERENCES

1. Espy-Wilson, C. Y. (**1992**). "Acoustic Measures for linguistic features distinguishing the semivowels in American English," J. Acoust. Soc. Am., **92**, pp. 736-757.

2. Nolan, F. (**1983**). *The phonetic bases of speaker recognition*, Cambridge University Press, Cambridge, England.

3. Lehiste, I. (**1962**). "Acoustical characteristics of selected English consonants," University of Michigan Communication Sciences Laboratory Report #9. Also

published in International Journal of American Linguistics (**1964**), **30**.

4. Peterson, G.E., & Barney, H.L., (**1952**). "Control methods used in a study of the vowels," J. Acoust. Soc. Am., **24**, pp. 175-184.
5. Daniloff, R. & Moll, K. (**1968**). "Coarticulation of lip-rounding," Journal of Speech and Hearing Research, **11**, pp. 707-721.
6. Keating, P. (**1988**). "Underspecification in Phonetics," Phonology **5**, pp. 275-292.
7. Perkell, J. S. & Matthies, M. L. (**1992**). "Temporal measures of anticipatory labial coarticulation for the vowel /u/: Within- and cross-subject variability," J. Acoust.l Soc. Am., **91**, pp. 2911-2925.
8. Bell-Berti, F. & Harris, K. S. (**1981**). "A temporal model of speech production," Phonetica **38**, 9-20.
9. Boyce, S. E., R. A. Krakow, F. Bell-Berti, and C. Gelfer. (**1990**). "Converging sources of evidence for dissecting articulation into core gestures," Journal of Phonetics **18**, pp. 173-188.
10. Browman, C. P. & Goldstein, L. M. (**1986**). "Towards an articulatory phonology," Phonology Yearbook **3**, pp. 215-252.
11. Bell-Berti, F. & Krakow, R. (**1991**). "Anticipatory velar lowering: A coproduction account," J. Acoust. Soc. Am., 90, 112-123.
12. Perkell, J. S., Cohen, M., Svirsky, M., Matthies, M., Garabieta, I. & Jackson, M. (**1992**). "Electro-Magnetic Midsagittal Articulometer (EMMA) systems for transducing speech articulatory movements," J. Acoust. Soc. Am., **92**, pp. 3078-3096.
13. Boyce, S. E. and Espy-Wilson, C. Y. (**submitted**). "Coarticulatory Stability of American English /r/," J. Acoust. Soc. Am.
14. Delattre, P. & Freeman, D. (**1968**). "A Dialect Study of American R's by X-ray Motion Picture," Language, **44**, pp. 29-68.

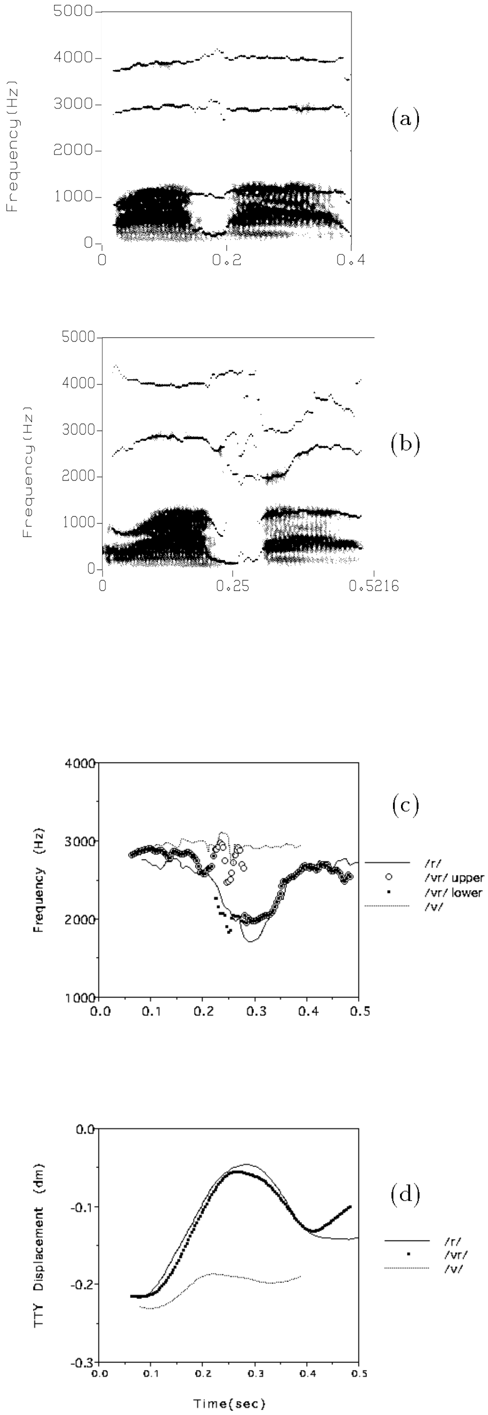

Figure 2: (a) Spectrogram of "wavav", (b) Spectrogram of "wavrav", (c) F3 tracks of /v/ in "wavav", /r/ in "wavrav" (upper and lower paths) and (d) TTY plots corresponding to F3 tracks in (c).