



ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

Speech Communication

Session 5aSCb: Production and Perception II: The Speech Segment (Poster Session)

5aSCb15. Articulatory overlap in English syllables with postvocalic /ɹ/

Rachel Walker* and Michael Proctor

*Corresponding author's address: Linguistics, University of Southern California, 3601 Watt Way, GFS 301, Los Angeles, CA 90089-1693, rwalker@usc.edu

In General American English (GAE), only two full vowels [ɑ, ɔ] occur in syllables ending in [ɹ] plus a non-coronal consonant, e.g. <harp>, <pork>. An articulatory study of rhotic production by three speakers of GAE was conducted using real-time structural Magnetic Resonance Imaging (rtMRI) (Narayanan et al. 2004). Subjects produced /ɹ/ in simple and complex syllable codas in a range of vocalic environments. Results show that the tongue dorsum shows the least movement in [-ɑɹ-] and [-ɔɹ-] sequences. This dorsal stability sheds light on why [ɑ] and [ɔ] are the only full vowels occurring before codas with /ɹ/ and a non-coronal consonant. English syllable rimes have been analyzed as maximally three timing units in length (Hammond 1999). Long vowels occupy two units, and most coda consonants occupy one unit, rendering rimes with long [ɑ]/[ɔ] followed by [ɹp] or [ɹk] problematic. We hypothesize that the high degree of overlap in the dorsal posture in [ɑɹ] and [ɔɹ] sequences allows their gestures to be partially blended and function like a diphthong that occupies two units in the rime. This study supports a view of maximal constituency in rimes with [ɹ] that takes articulatory overlap into account.

Published by the Acoustical Society of America through the American Institute of Physics

INTRODUCTION

English syllables containing liquid consonants exhibit metrical and phonotactic properties that can provide important insights into syllable structure, but they are not yet well understood. Of particular interest are liquids in the rime, because of the restrictions on other segments that can appear in the same syllable, the ways in which the liquids interact with their tautosyllabic segments, and the implications for syllable weight. In this study, we examine some sequencing constraints involving vowel-rhotic-consonant interactions in American English rimes, and we demonstrate, using MRI, some factors in the articulation of these syllables which bear upon these phenomena. While many of the data we address are also relevant to other varieties of English, here we focus our attention on General American English (GAE): rhotic American varieties typified by the English spoken throughout much of the Midwestern United States (Wells, 1982).

BACKGROUND

There is consensus across several studies that syllable rimes in GAE make available up to three units, such as abstract temporal slots or morae, which provide an upper limit on the rime's constituency (Fujimura, 1979; Borowsky, 1989; Hammond, 1999; Hall, 2001). Long (tense) vowels and diphthongs typically occupy two units, and each short (lax) vowel or consonant occupies one unit. Maximal rimes may consist of a long vowel or diphthong followed by a single coda consonant (e.g. *pike* [pa:k], *peek* [pi:k], *loom* [lu:m]), or a short vowel followed by two coda consonants (e.g. *hasp* [hæsp], *milk* [mɪlk], *rump* [ɹʌmp]). Rimes that exceed the three-unit upper limit are not usually possible in GAE; however, rimes can be followed by an appendix of coronal obstruents, as in *text* [tekst], *traipse* [tɹeɪps], and *lounge* [laundʒ] (Fujimura & Lovins, 1978; Fujimura, 1979; Halle & Vergnaud, 1980; Kiparsky, 1980; Selkirk, 1982; Borowsky, 1989; Hall, 2001).

Despite the broad-based conformity of GAE rimes with the three-unit maximum, it is challenging to reconcile rimes with post-vocalic [ɹ] with this generalization. Long [ɑ] and [ɔ] are the only full vowels that occur before codas containing [ɹ] plus a non-coronal consonant, as in words such as *harp* and *pork*. Words like these appear to exceed the rime size limit. The long status of [ɑ] and [ɔ] in GAE is diagnosed by their capacity to occur in vowel-final monosyllabic words (e.g. *spa* [spɑ:], *paw* [pɔ:]) (Lass, 1976; Hammond, 1997, 1999; Ladefoged & Johnson, 2011). If each coda consonant occupied a rime unit (apart from coronal obstruent appendices) in addition to the long vowel, words like *harp* and *pork* would contain rimes with four units, corresponding to the transcriptions [ha:ɹp] and [pɔ:ɹk].

In this study, we investigated two interconnected questions. First, why are [ɑ] and [ɔ] the only full vowel qualities attested before a coda consisting of [ɹ] and a non-coronal consonant in GAE? Second, how can rimes consisting of these sequences be reconciled with the apparent restriction of GAE rimes to a three-unit maximum?

This research pursues the possibility that the special distribution of [ɹ] in GAE rimes derives from the vowel-like dorsal constriction that it shows in addition to its coronal constriction (Delattre & Freeman, 1968; Zawadzki & Kuehn, 1980). We hypothesize that it is not the generic long or short character of the vowel before [ɹ] that is relevant, but rather the degree of coincidence of the tongue dorsum posture during the vowel and [ɹ]. In particular, the high stability of the dorsal posture during [-ɑɹ-] and [-ɔɹ-] sequences allows their gestures to be partially blended and function like a complex vocoid that occupies two units in the rime, where the production of intrinsic gestures of the vowel and rhotic overlap in time to a significant extent.

REAL-TIME MRI STUDY

An articulatory study of rhotic production by three native speakers of GAE was conducted using real-time structural Magnetic Resonance Imaging (rtMRI) (Narayanan et al., 2004). Subjects produced [ɹ] in simple and complex syllable codas in a range of vocalic environments. The rtMRI technique captures images of the entire vocal tract in the midsagittal plane at 33 frames/sec, allowing us to examine the constriction locations for [ɹ] and a preceding vowel and the dynamics of their production.

Method

Stimuli used for this study are given in Table 1.¹ Each target word was elicited three times in a short carrier phrase constructed from non-lingual consonants and symmetrical vowels, e.g. *hee beer hee, ha barb ha*. Not all speakers of GAE produce [uɹ-] sequences. One subject pronounced *moor* and *cured* with [u].

TABLE 1. Elicitation items: monosyllabic words with [ɹ] in simple coda (center column) and coda-cluster (right column) contexts.

Nucleus	-I#	-IC#
[i]	<i>beer</i>	<i>beard</i>
[u]	<i>moor</i>	<i>cured</i>
[ɔ]	<i>bore</i>	<i>form</i>
[ɑ]	<i>bar</i>	<i>barb</i>

MRI data were acquired at Los Angeles County Hospital on a Signa Excite HD 1.5T scanner. Subjects' upper airways were imaged while they lay supine in an MRI scanner. Target words were presented in carrier phrases one at a time, at a natural speaking rate. A custom pulse sequence was used to acquire a 5 mm midsagittal slice with image resolution of 68×68 pixels over a 200×200 mm field of view (Fig. 1). Image data were reconstructed as a 33.2 frame/sec video (Bresch et al., 2008), and noise-reduced speech audio was simultaneously recorded at a sampling frequency of 20 kHz (Bresch et al., 2006).

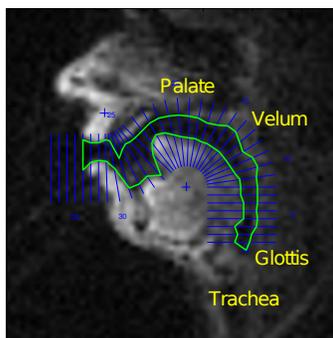


FIGURE 1. Real-time Magnetic Resonance Image frame captured during articulation of [ɹ]. Analysis grid (blue lines) and vocal tract outline generated by semi-automatic segmentation of air-tissue boundaries (green lines) superimposed on MR image.

Analysis

For each utterance, the articulatory midpoint of [ɹ] was estimated by locating the frame with maximal lingual elongation between coronal and pharyngeal constrictions. A 300 to 350 ms interval centered on that frame was identified, corresponding to the interval of production for [ɹ]. Start and end points for each interval were chosen to be the articulatory midpoints of neighboring vowels, which were estimated by locating the frame where the tongue was maximally static in the vicinity of the target constriction for the vowel.

For each frame, the tongue posture in the midsagittal plane was captured by automatically identifying air-tissue boundaries (Proctor et al., 2010), and manually correcting the tongue outline where the algorithm failed to locate the edges of lingual tissue with sufficient accuracy (Fig. 1). In subsequent figures, air-tissue boundaries like those identified by the green lines in Figure 1 are displayed without the original magnetic resonance image. The chronological sequence of air-tissue boundaries across frames depicts the course of articulation in the midsagittal plane, tracking tongue movement between the vowel and consonant of interest (Fig. 2). Because subjects' heads remained stationary throughout data acquisition, tongue position can be compared across tokens for each subject.

¹ Target words with the rhotic vowel (*fur, firm*) were also collected, but we do not report on them here.

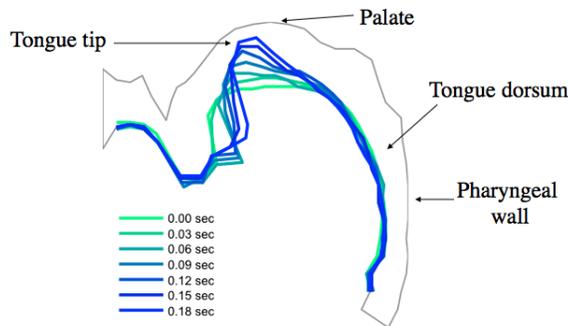


FIGURE 2. Analysis of articulatory movement in rtMRI data for sequence [-aɪ#]. Each superimposed curve represents the location of the tongue edge in the midsagittal plane at a given point in time. Lingual posture is tracked at 30 ms increments, in the interval from the midpoint of the nuclear vowel (light green) to the midpoint of post-vocalic [ɪ] (dark blue).

Results

The data reveal different patterns of tongue movement in the production of vowel-rhotic sequences by these speakers of GAE. Syllable-final rhotics produced after high front vowels demonstrate more tongue root movement into the syllable coda (Fig. 3 left) compared with coda rhotics produced after back vowels (Fig. 3 center and right), where the tongue root remains more stable throughout production of the entire rime. Furthermore, in the [ɔɪ] and [aɪ] sequences produced by the three subjects examined here, the posture assumed by the tongue dorsum and tongue root during the nucleus more closely resembled the intrinsic dorsal posture observed during rhotic production (Fig. 3 right).² Overall similar patterns of articulation were observed in the production of stimuli with complex codas.

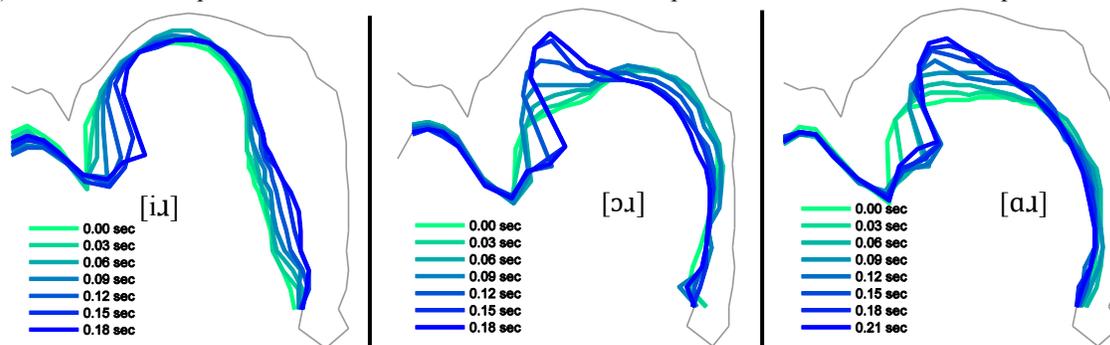


FIGURE 3. Articulation of vowel-rhotic sequences by one speaker of GAE: tongue movement into rhotic codas following a high front vowel (left panel), a mid-back vowel (center panel), and a low vowel (right panel). Greater tongue root stability can be observed in rhotic production after non-front vowels [ɔ] and [a], where the dorsal posture of the vowel more closely matches that of the intrinsic tongue body gesture of the rhotic.

IMPLICATIONS FOR RIME STRUCTURE

Given the similarities of the dorsal postures for the non-high back vowels [a] and [ɔ] and the rhotic [ɪ], we propose that the component intrinsic tongue body gestures in GAE [-aɪ-] and [-ɔɪ-] sequences are partially blended, giving rise to a complex vocoid. In parity with long monophthongs or diphthongs, [aɪ] and [ɔɪ] occupy just two weight units in the rime. As a result, rimes in words like *harp* and *pork* amount to three units, two for the partly coalesced vowel-[ɪ] sequence, and one for the final non-coronal consonant, corresponding to revised transcriptions [hɑɪp] and [pɔɪk]. The vowels are transcribed as short here, representing the portion that is not co-produced with [ɪ]. The remainder of the vowel is realized simultaneously with the dorsal and root posture for the rhotic. These rimes are comparable in size to those in words like *pike* [paɪk] and *hype* [hɑɪp], where the final coda consonant contributes one unit and [aɪ] contributes two.

² The subject whose productions are represented in Fig. 3 showed this effect most noticeably in [aɪ] sequences. Even greater dorsal stability was observed during production of [ɔɪ] sequences by the other subjects.

Our study found that the tongue dorsum and root posture of GAE vowels other than [ɑ] and [ɔ] show a greater departure from that of [ɪ], a finding verified with a broader range of vowel qualities by Proctor and Walker (2012). Because of the greater difference in articulation, we infer that in a sequence like [-i:ɪ-], the vowel and rhotic do not share intrinsic dorsal gestures, so the sequence is structured as a long vowel followed by coda [ɪ], representing a sequence of distinct dorsal targets, rather than as a partially blended complex vocoid. A rime with a sequence like [-i:k] is thus not attested in GAE, because it will exceed the three-unit maximum, with [i:] contributing two units and each of [ɪ] and [k] contributing one.

The structure that we propose for [-ɑɪ-] and [-ɔɪ-] sequences in rimes finds precursors in previous proposals that post-vocalic [ɪ] can form a diphthong with a tautosyllabic vowel in GAE (Harris, 1994; Green, 2001). However, our analysis points to important structural differences between combinations of specific vowels and [ɪ] in GAE rimes, since the nature of vocoid that emerges in these sequences depends on the properties of both nuclear vowel and rhotic elements; in particular, where the nuclear vowel shows sufficient similarity to [ɪ] in its defining dorsal gesture, it has the capacity to form a different type of partially blended sequence than non-back vowels.

Returning to our questions at the outset, we asked why [ɑ] and [ɔ] are the only full vowels attested in rimes containing [ɪ] plus a non-coronal consonant, and how rimes consisting of these sequences are reconciled with the three-unit maximum for GAE rimes. The answers we have proposed are intertwined. We suggest that only [ɑ] and [ɔ] are attested because these are the only vowels that can form an extended complex vocoid with [ɪ] that occupies two units in the rime. The two-unit size of this partially blended structure also explains how rimes with [-ɑɪ-] and [-ɔɪ-] sequences followed by a non-coronal consonant conform to the maximal size limit.

ACKNOWLEDGMENTS

This work was supported by National Institutes of Health grants NIDCD-RO1-DC-006705 and NIDCD-RO1-DC-002717. For comments on this work, special thanks to Steve Parker, Ruben van de Vijver, the Speech Production and Articulation Knowledge (SPAN) group, the phonetics and phonology groups at the University of Southern California, and audiences at the University of Toronto, the University of Melbourne, and OCP 9.

REFERENCES

- Borowsky, T. (1989). "Structure preservation and the syllable coda in English," *Natural Language and Ling. Theory* 7, 144-166.
- Bresch, E., Kim, Y.-C., Nayak, K., Byrd, D., and Narayanan, S. (2008). "Seeing speech: Capturing vocal tract shaping using real-time magnetic resonance imaging [Exploratory DSP]," *Signal Processing Magazine, IEEE Sig. Proc. Mag.* 25, 123-132.
- Bresch, E., Nielsen, J., Nayak, K., and Narayanan, S. (2006). "Synchronized and noise-robust audio recordings during realtime MRI scans," *J. Acoust. Soc. Am.* 120, 1791-1794.
- Delattre, P., and Freeman, D. C. (1968). "A dialect study of American r's by x-ray motion picture," *Linguistics* 44, 29-68.
- Fujimura, O. (1979). "An analysis of English syllables as cores and affixes," *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 32, 471-476.
- Fujimura, O., and Lovins, J. (1978). "Syllables as concatenative phonetic units," in *Syllables and Segments*, edited by A. Bell and J. B. Hooper (North Holland, Amsterdam), pp. 315-334.
- Green, A. D. (2001). "American English 'r-colored' vowels as complex segments," *Linguistics in Potsdam* 1, 70-78.
- Hall, T. A. (2001). "The distribution of superheavy syllables in Modern English," *Folia Linguistica* XXXV/3-4, 399-442.
- Halle, M., and Vergnaud, J.-R. (1980). "Three-dimensional phonology," *Journal of Linguistic Research* 1, 83-105.
- Hammond, M. (1997). "Vowel quantity and syllabification in English," *Language* 73, 1-17.
- Hammond, M. (1999). *The Phonology of English: A Prosodic Optimality-Theoretic Approach*, (Oxford UP, Oxford).
- Harris, J. (1994). *English Sound Structure*, (Blackwell, Oxford).
- Kiparsky, P. (1980). "Remarks on the metrical structure of the syllable," *Phonologica* 1980, 245-256.
- Ladefoged, P., and Johnson, K. (2011). *A Course in Phonetics*, 6th Ed., (Wadsworth, Boston).
- Lass, R. (1976). *English Phonology and Phonological Theory: Synchronic and Diachronic Studies*, (Cambridge UP, Cambridge).
- Narayanan, S., Nayak, K., Lee, S., Sethy, A., and Byrd, D. (2004). "An approach to real-time magnetic resonance imaging for speech production," *J. Acoust. Soc. Am.* 115, 1771-1776.
- Proctor, M. I., Bone, D., and Narayanan, S. S. (2010). "Rapid semi-automatic segmentation of rtMRI for parametric vocal tract analysis," *Proceedings of InterSpeech, Makuhari, Japan*.
- Proctor, M., and Walker, R. (2012). "Articulatory bases of sonority in English liquids," in *The Sonority Controversy*, edited by S. Parker (Mouton de Gruyter, Berlin), pp. 289-316.
- Selkirk, E. O. (1982). "The syllable," in *The Structure of Phonological Representations*, edited by H. van der Hulst and N. Smith (Foris Publications, Dordrecht), pp. 337-383.
- Wells, J. C. (1982). *Accents of English 3: Beyond the British Isles*, (Cambridge UP, Cambridge).
- Zawadzki, P. A., and D. P. Kuehn. (1980). "A cineradiographic study of static and dynamic aspects of American English /r/," *Phonetica* 37, 253-266.