

Intergestural Timing in English /r/

Bryan Gick^{†‡} and Fiona Campbell[†]

[†] University of British Columbia, Canada

[‡] Haskins Labs, New Haven CT, USA

E-mail: gick@interchange.ubc.ca, fionaca@interchange.ubc.ca

ABSTRACT

Relative timing of the gestures of /r/ was measured in initial, final, and intervocalic positions for eight speakers representing several dialects of English. Simultaneous video and ultrasound were used to collect kinematic midsagittal measurements of several events associated with English /r/. Results show that, in initial allophones, gestures occur front-to-back, with the lip gesture occurring first, then the tongue body, then the tongue root. In final position, there is little timing effect, but spatially, the lip gesture is often obscured or reduced. Timing is not significantly affected by resyllabification or dialect. These results show timing patterns for /r/ that are analogous in crucial ways to those previously observed for /l/ and /w/, but that further support that these timing differences are phonetic responses to perceptual recoverability factors. An additional observation of this paper is that tongue mid lowering is likely a tongue body raising mechanism.

1. INTRODUCTION

English /r/ has been the focus of a wide range and variety of previous work, including studies of the articulatory complexity of /r/, cross-dialect and cross-speaker differences in /r/ production, phonological phenomena unique to /r/, and so on [e.g., 1, 2]. While recent work in relative timing of gestures has provided an articulatory basis for describing the syllable and a wide range of syllable-based phenomena [3, 4, 5, 2, 6], no previous studies have measured intergestural timing of all of the gestures of English /r/. In most dialects of English, /r/ comprises three independent oral constrictions: tongue root (TR), tongue tip/body (TB) and lips (Lip), making it one of the most articulatorily complex segments of English. The goal of the present study is to discover how this complexity is realized in terms of intergestural timing. Understanding the relative timing of these three gestures will thus not only shed light on controversies surrounding English /r/, as it has done for other English segments such as /l/ and nasals, but will also help to disambiguate previous generalizations based on observations of interactions between only two gestures.

It has previously been observed that certain gestures (e.g., the anterior closure for /l/ or nasals) tend to occur more peripherally in syllables while other gestures (e.g., the tongue dorsum gesture for /l/ or the velum gesture for nasals) occur more medially. By way of explaining this

observation, earlier studies [4, 5] suggested that the more peripheral gestures were 'consonantal' and more central ones 'vocalic', as determined by their degrees of constriction. Later work [7] found similar temporal effects in /w/, showing that it is not degree but relative location (specifically anteriority) of constriction that determines relative timing. This view lends support to the proposal that this front-to-back ordering is driven by perceptual recoverability [6], whereby more anterior gestures are known to mask less anterior ones, forcing a timing offset. However, because all previous studies have looked only at timing of two gestures at a time, the question remains unanswered as to whether this response (if it is driven by recoverability) is a phonetic response to perception factors, or a phonologized timing pattern, truly representative of the structure of the syllable in terms of gestures.

The present paper reports the results of an experiment measuring the relative timing of the gestures of /r/ for eight speakers representing a number of dialects of English. If the observed timing pattern is a phonetic response to perceptual factors, we should expect to see a gradient response to anteriority (i.e., the most anterior gesture should peak first [in onset position], followed by the second-most anterior gesture, followed by the least anterior gesture); if, rather, the surface timing pattern is the result of a phonological categorization of gestures, whether driven historically by perceptual factors or not, we should expect to see a two-way distinction between the behavior of more consonant-like gestures versus more vowel-like ones (with more anterior gestures possibly classifying as 'consonantal'). The latter possibility is further supported by the limited previous observations of spatial reduction/augmentation, where the two more anterior gestures of /r/ have been found to follow typically consonant-like patterns of reduction (i.e., reduction in final allophones) while the tongue root gesture shows a categorically opposite effect (augmentation in final position) [2, 8].

In addition to measuring the three gestures necessary for testing the above hypotheses, a fourth event – the timing of the lowering of the tongue mid (TM; corresponding to a constriction of the middle genioglossus muscle) – were also measured. Testing whether TM lowering correlates with one of the other two lingual gestures will help in understanding the physical mechanism for producing English /r/. As MacNeilage [9] puts it (p. 372): "But what of the contraction of the middle third of the genioglossus

which allows us to put a dent in the tongue between the two constrictions...If not a gesture in itself, which of the other two gestures is it a part of, or is it to be considered a part of both?" Not only have previous studies questioned this mechanism in general, but other studies [4] have assumed an analogous correspondence between tongue mid lowering and tongue dorsum backing for /l/. A clear correlation may allow for easier measurement of /r/ gestures in future studies.

2. METHODS

An experiment was conducted using video and ultrasound to image the three gestures of /r/ in pre-, inter-, and postvocalic positions.

2.1. Subjects

Eight native speakers of English participated in the study. Speakers of the following dialects were included: Western Canadian (WC; 2 speakers), Wisconsin (WI; 2 speakers), Massachusetts (MA; 2 speakers, rhotic dialect), Upper Peninsula Michigan (MI; 1 speaker), Southeastern British Received Pronunciation (RP; 1 speaker). All subjects were students living in Vancouver at the time of data collection, and all were between the ages of 20 and 28. Data from the MI speaker was not included in the analysis due to the lack of measurable lip movement present for this speaker. As expected, only initial /r/ was present in the dialect of the RP speaker. While not enough measurements could be made to statistically analyze this subject's data separately, the RP initial /r/ was included in the combined cross-dialect analysis as they exhibited a pattern similar to that of the other dialects.

2.2 Stimuli

Stimuli were modeled after previous studies of /l/ and /w/ (e.g., [7]), with /r/ flanked by maximally similar vowels, and only the syllable/word position of the /r/ changing across conditions, thus:

initial r	intervocalic r	final r
...V ₁ #RV ₁V ₁ R#V ₁V ₁ R#hV ₁ ...

Because the three gestures of /r/ span the entire length of the vocal tract, there is no vowel context that does not interfere with some gesture. Thus, two contexts were used: /a/ and /i/ were chosen as V₁ because 1) they are both unrounded and therefore do not obscure the labial gesture, which can then be used as a baseline, and 2) these vowels each interfere with only one of the lingual gestures of /r/ allowing observation of the other. /h/ was included in the Final condition to prevent resyllabification. A nonsense phrase containing the target context and /r/ in a carrier sentence was read 10 times (in random order) by each subject. Example: *He said "har haw" for me.*

2.3 Procedure

Lingual data were recorded to DV tape at standard video

rate (29.97 frames/sec) from an Aloka SSD-900 portable ultrasound machine. A 3.5MHz convex intercostal probe was fixed to a microphone stand and held against each subject's neck, just above the larynx, so that a midsagittal section of the tongue was visible from root to tip [10]. Lips were colored red, and glasses with a red marker were worn; lip and head position data were recorded using a video camera, and the reds were automatically extracted from the image. The extracted reds were overlaid on the ultrasound image frame-by-frame using a video mixer.

2.4 Analysis

One judge, uninformed about the nature of the experiment, but trained in ultrasound video analysis was instructed to record frame numbers in which subjects appeared to have reached the target for each of Lip, TB, TM, and TR for each instance of /r/. See figure 1 for an example of approximate measurement locations. Recall that, as every vowel of English interferes with at least one gesture of /r/, it was impossible in most cases to make direct comparisons between all of the relevant /r/ gestures for any one instance of /r/. Because of this the labial gesture (which is visible in both /i_i/ and /a_a/ contexts) was used as a baseline for comparison with the TB raising (visible in the /a_a/ context) and TR backing (visible in the /i_i/ context). By subtracting the times of each of the other two gestures from that of the Lip maximum it is possible to ascertain the relative timing of each event associated with /r/ in relation to the others.

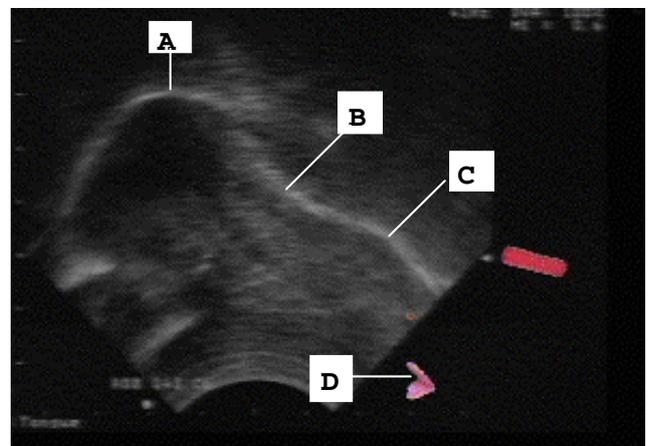


Figure 1: Approximate measurement locations for a) TB raising, b) TM lowering, c) TR retraction, and d) Lip rounding/protrusion.

3. RESULTS

3.1. Intergestural timing lag

Resyllabification had no significant effect on timing (i.e., there was no significant difference between word-final /r/'s when they preceded vowel-initial vs. consonant-initial words). This is consistent with previous results for /l/ and /w/ [7]. Thus, all lexically word-final /r/'s will be combined for the present analysis.

Figure 2 shows a small, though significant, timing

difference between initial and final allophones for TB [$F(1,129) = 12.008, p < .001$] and TR [$F(1,160) = 34.758, p < .0001$] gestures. In both cases, there is a positive lag relative to Lip (i.e., Lip precedes each gesture) in the initial allophone, and a smaller negative lag (i.e., Lip precedes each gesture) in the final allophone. Thus, in the initial allophone, Lip occurs about 10ms earlier than TB, and about 20ms earlier than TR. By subtracting these ((TR-Lip)-(TB-Lip)), it can be seen that there is a very small similar difference of about 10ms between TB and TR in initial position. For the final allophone, the mean difference between all three gestures is well under 10ms. These should be considered to be essentially simultaneous.

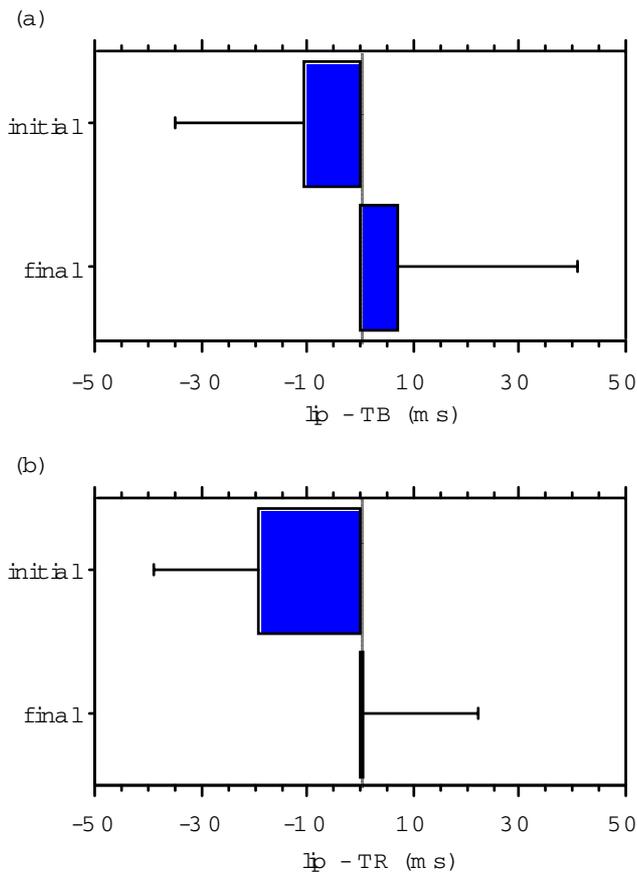


Figure 2: a) TB and b) TR lag, relative to Lip.

There was no qualitative differences in timing between dialects. WI speakers showed the pattern most prominently, with a slightly greater magnitude of effect, but the general pattern was still present for other dialects.

3.2. Spatial reduction of lip gesture

The only substantial difference between dialects – and the only visible effect of resyllabification – was in reduction of the Lip gesture. Because magnitudes were not quantified in the present study, only complete reduction (i.e., reduction to zero) was measurable (calculated as the percent of tokens having a measurable Lip event). For this measure, all three dialects showed a marked difference between lexically initial and final /r/. Only WI showed a

strong effect of resyllabification (i.e., a substantial increase in the presence of a measurable Lip gesture when final /r/ is followed by a vowel-initial word). MA showed the least effect of final reduction at 42%, while WC showed an almost complete loss of the Lip gesture in final allophones.

Dialect	ha#raw	har#awe	har#haw
WI	100	70	17
MA	100	54	42
WC	100	4	0
Mean	100	43	20

Table 1: Percent of tokens with a measurable lip gesture.

3.3. Timing of tongue mid lowering.

Figure 3 shows the relationship of TM lowering to a) TB raising and b) TR backing. A paired t-test showed that while TR was found to be significantly different from TM ($df = 122; p < .001$), there is no difference between TM and TB ($df = 125; p = .5795$). This strongly suggests that TM lowering is a mechanism for TB raising rather than for TR backing or for both.

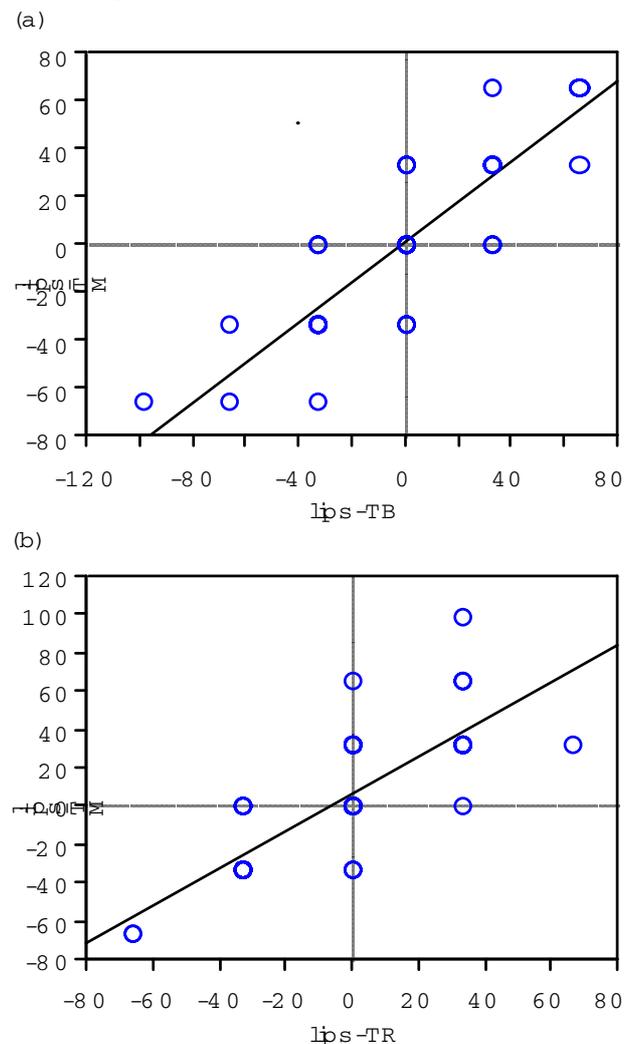


Figure 3: Scatterplots of a) TM * TB and b) TM * TR.

4. DISCUSSION

The results of this study have shown:

1) Across dialects, gestures occur front-to-back in initial allophones of /r/, with the lip gesture peaking first, then the tongue body, then the tongue root. This is contrary to a phonological view of these patterns, and supports the hypothesis that allophonic patterns in articulatory timing are gradient phonetic responses to the need to maintain perceptual recoverability of all of the gestures of a segment. There is essentially no timing offset between /r/ gestures in final and intervocalic positions, and no substantial differences between dialects, further suggesting that this temporal offset may be driven by more universal phonetic factors. Thus, the perceptual recoverability model is supported in this study.

2) Spatially, the Lip gesture is largely obscured or reduced word-finally in all dialects. Unlike temporal offset, however, this reduction is affected by both resyllabification and dialect. Spatial characteristics thus appear to behave more like language-specific phonological properties. The results in 1 and 2 are analogous in crucial ways to patterns previously observed for /l/ and /w/.

3) Finally, tongue mid lowering was found to pattern closely with the more anterior tongue body gesture rather than with the tongue root gesture, suggesting that the tongue mid lowering is likely a mechanism to facilitate the tongue body raising gesture.

ACKNOWLEDGEMENTS

Many thanks to Julee Botting, Marie Jette, Lisa Loo, Carla Monteleone, Sunyoung Oh, Sharon Park, and Shaffiq Rahemtulla for assistance with data collection and analysis. Special thanks to Ramona McDowell for analysis work. This research was supported by an NSERC individual grant to the first author, and NIH grant DC-02717 to Haskins Laboratories.

REFERENCES

- [1] P. C. Delattre and D. C. Freeman. "A dialect study of American r's by x-ray motion picture," *Linguistics*, vol. 44, pp. 29-68, 1968.
- [2] B. Gick, "A gesture-based account of intrusive consonants in English," *Phonology*, vol. 16, pp. 29-54, 1999.
- [3] R. Krakow. *The Articulatory Organization of Syllables: A Kinematic Analysis of Labial and Velar Gestures*, PhD dissertation, Yale University, 1989.
- [4] R. Sproat and O. Fujimura, "Allophonic variation in English /l/ and its implications for phonetic implementation," *JPhon*, vol. 21, pp. 291-311. 1993.
- [5] C. P. Browman and L. Goldstein, "Gestural Syllable

Position Effects in American English," in *Producing speech: contemporary issues. For Katherine Safford Harris*, F. Bell-Berti and L. Raphael, Eds., pp. 19-34. Woodbury NY: American Institute of Physics, 1995.

- [6] A. Kochetov, *Production, Perception, and Emergent Phonotactic Patterns: A Case of Contrastive Palatalization*. New York & London: Routledge. 2002.
- [7] B. Gick, "Articulatory correlates of ambisyllabicity in English glides and liquids," in *LabPhonVI: Constraints on Phonetic Interpretation*, J. Local, R. Ogden and R. Temple, Eds., Cambridge: Cambridge U. Press, In press.
- [8] B. Gick and L. Goldstein, "Relative timing of the three gestures of North American English /r/," *JASA*, vol. 111(5), Pt. 2, p. 2481, 2002.
- [9] P. MacNeilage, "The relation between speech perception and speech production in phylogeny and ontogeny," in *CLS33: The Panels*, pp. 369-379.
- [10] B. Gick, "The use of ultrasound for linguistic phonetic fieldwork," *JIPA*, In press.