Spatial and Temporal Properties of Gestures in North American English /ɪ/

Fiona Campbell[1], Bryan Gick[1,2], Ian Wilson[3], Eric Vatikiotis-Bateson[1,2]

[1] *University of British Columbia, Vancouver, BC, Canada*
[2] *Haskins Laboratories, New Haven, Conneticut, U.S.A*
[3] *University of Aizu, Japan*

Key words

articulatory gestures

articulatory timing

English /ɪ/

Optotrak

ultrasound

Abstract

Systematic syllable-based variation has been observed in the relative spatial and temporal properties of supralaryngeal gestures in a number of complex segments. Generally, more anterior gestures tend to appear at syllable peripheries while less anterior gestures occur closer to syllable peaks. Because previous studies compared only two gestures, it is not clear how to characterize the gestures, nor whether timing offsets are categorical or gradient. North American English /r/ is an unusually complex segment, having three supralaryngeal constrictions, but technological limitations have hindered simultaneous study of all three. A novel combination of M-mode ultrasound and optical tracking was used to measure gestural relations in productions of /r/ by nine speakers of Canadian English.

Results show a front-to-back timing pattern in syllable-initial position: Lip then tongue blade (TB), then tongue root (TR). In syllable-final position TR and Lip are followed by TB. There was also a reduction in magnitude affecting Lip and TB gestures in syllable-final position and TR in syllable-initial position. These findings are not wholly consistent with any theory advanced thus far to explain syllable-based allophonic variation. It is proposed that the relative magnitude of gestures is a better predictor of timing than relative anteriority or an assigned phonological classification.

Acknowledgements: The authors would like to thank Patricia Shaw, Sonya Bird, and Douglas Whalen for their input and helpful comments, and Shaffiq Rahemtulla and Nahal Namdaran for assistance with equipment, data collection, and analysis. This research was funded by an NSERC Discovery Grant to the second author and NIH Grant DC-02717 to Haskins Laboratories.

Address for correspondence. Fiona Campbell, Department of Linguistics, 1085 Dr. Penfield, McGill University, Montreal, Quebec, H3A 1A7 Canada; e-mail <fionaca@gmail.com>

Language and Speech 0023-8309; Vol 53(1): 49–69; 351209; DOI:10.1177/0023830909351209
http://las.sagepub.com

1 Introduction

Previous work on syllable-based allophonic variation has shown that the relative timing of the two oral gestures in English /l/, /w/, and nasals is such that the more anterior gestures (those occurring physically farther forward in the vocal tract) tend to appear at syllable peripheries (see Krakow, 1999, and Gick, Campbell, Oh, & Tamburri-Watt, 2006, for recent summaries of this literature). In addition, allophonic variation of complex segments has been linked to position-dependent spatial reduction of gestures (e.g., Sproat & Fujimura, 1993), such that the less anterior gesture of the two gestures tends to have a smaller magnitude in syllable-initial position, and the more anterior gesture shows a reduction in magnitude in syllable-final position. It is not clear what drives these phenomena, particularly whether they are categorical effects encoded in a speaker's phonology or phonetic effects resulting from perceptual or biomechanical factors. Additionally, while timing and magnitude patterns have often been examined in tandem, it is not clear whether these are independent or linked. The goal of the present article is to test hypotheses drawn from previous studies of two-gesture segments by examining the syllable-based variation in timing and magnitude of gestures in a segment with three supralaryngeal gestures: North American English /r/.

1.1 Background

As it is produced in most North American dialects of English, /r/ is unusually complex (e.g., Alwan, Narayanan, & Haker, 1997; Delattre & Freeman, 1968; Docherty & Foulkes, 2001; Espy-Wilson, 2004; Guenther, Espy-Wilson, Boyce, Matthies, Zandipour, et al., 1999; Hagiwara, 1995; Hashi, Honda, & Westbury, 2003; Lindau, 1985; Uldall, 1958; Tiede, Boyce, Holland, & Choe, 2004; Westbury, Hashi, & Lindstrom, 1998) in that, although the exact lingual configuration varies widely, it is generally composed of three independent supralaryngeal constrictions: one between the tongue root and the pharyngeal wall (TR), one between the tongue tip/body and the palate or the alveolar ridge (TB), and one between the lips (Lip). This complexity makes /r/ uniquely suitable for testing whether the observed gestural timing and magnitude patterns are gradient (likely phonetic) or categorical (likely phonological) effects. However, the difficulty of imaging and measuring movements in the lip, hard palate, and pharyngeal regions simultaneously during speech has impeded such study.

Position-dependent variation in the magnitude of the gestures in /r/ has been reported in a number of studies. Delattre and Freeman (1968) used cineradiograms (x-ray films) to document cross-dialectal and cross-subject variation in North American English /r/. They noted that lip rounding and the retroflex tongue shape (or the closest thing to it) are more likely to occur in a strong syllabic position (e.g., prevocalic pre-stress). Zawadzki and Kuehn (1980), also using cineradiograms, observed variation in tongue shape and a difference between prevocalic and postvocalic allophones: "the prevocalic allophone was characterized by greater lip rounding, a more advanced tongue position, and less tongue dorsum grooving" (p.253). Gick (1999) used an electro-magnetic midsagittal articulometer (EMMA) to look at the magnitude of the more anterior lingual gesture (tongue tip or tongue blade) across positions and found a reduction in syllable-final allophones. Based on a simple probe-contact experiment, Hagiwara (1995) found that while "tip up" (retroflex) was a stable tongue shape across positions for a given speaker, subjects who used a "blade up" configuration in syllable-initial position were likely to use a different "tip down" configuration in syllable-final position. In a Magnetic Resonance Imaging (MRI) study, Alwan et al. (1997) found no positional differences between syllable-initial and syllabic /r/, but this is likely due to the fact that MRI requires sustained productions that obscure differences observed in continuous speech.

1.2 Predictions

Previous studies of two-gesture segments provide, either directly or indirectly, different predictions for timing between the three gestures of English /r/.

Krakow (1989) found that the velum lowering gesture for /m/ preceded the lip gesture in syllable-final position and followed it in syllable-initial position; also, the velum gesture was larger and longer in syllable-final position (regardless of stress pattern) than in syllable-initial position. In a 1993 study of North American English /l/, Sproat and Fujimura found that the tongue tip gesture preceded the tongue dorsum gesture and had a greater magnitude in syllable-initial position, and that the tongue dorsum gesture preceded the tongue tip gesture and had a greater magnitude in syllable-final position. To account for this (and for Krakow's observations of timing in nasals) they proposed that, based on the width of the constriction, a gesture could be classified as either intrinsically [consonantal] (producing "an extreme obstruction in the vocal tract" [1993, p.304]) or intrinsically [vocalic] (producing a less extreme obstruction, or an opening, as with the velum). In their view, the timing pattern observed was due to an attraction of vocalic gestures to the syllable nucleus, and of consonantal gestures to syllable margins. They add that consonantal gestures are "stronger" (i.e., have greater magnitude) in syllable-initial position and "weaker" (i.e., have less magnitude) in syllable-final position while vocalic gestures show the opposite pattern.

The present article assumes the three /r/ gestures to be [vocalic] in Sproat and Fujimura's model, as all three result in approximate constrictions in all syllable positions. Within this model, each gesture should be attracted to the nucleus to an equal degree, that is, all three gestures should be essentially simultaneous in both prevocalic and postvocalic positions and should pattern together in terms of the "strength" of the gesture across syllable positions. It may be, however, that the TB gesture is more appropriately categorized as [consonantal] for at least some speakers. Using Alwan et al.'s (1997) MRI data from two speakers of American English, Espy-Wilson, Boyce, Jackson, Narayanan and Alwan (2000) found that, while one speaker exhibited similar constriction areas for all three /r/ gestures, the other showed a considerably tighter constriction in the palatal region (see Figure 4, p.347). Both of these possibilities will be considered in the course of this article.

Browman and Goldstein's (1995) results for American English /l/ were similar to those of Krakow (1989) and Sproat and Fujimura (1993), except that the two gestures studied tended toward simultaneity in syllable-initial position. This result is consistent with their earlier proposal (Browman & Goldstein, 1992) that there is a "single syllable-final organizational pattern in which the wider constrictions always precede the narrower constrictions" (p.167), thus linking magnitude and timing in a gradient relationship (in syllable-final position only). In addition, they observed syllable-final reduction of the tongue tip gesture in /t/, /n/, and /l/ (the more anterior and more "consonant-like" gesture), as compared to syllable-initial position, and called this a “general positional effect” (note that an alternate view, in which the effect is syllable-initial augmentation, is also possible: Fougeron & Keating, 1997). As the size of a gesture is known to vary with syllable position, this view may be interpreted as allowing for a complex relationship between gestural magnitude and timing. For example, if the TB gesture of /r/ is found to have a greater magnitude than the TR gesture, it should follow the TR gesture in syllable-final position. Under this view (which is not explicitly included in Browman & Goldstein, 1995), this study predicts that the order of /r/ gestures in syllable-final position should be dependent on the relative magnitude of each gesture, while in syllable-initial position, /r/ gestures should tend toward simultaneity.

In a study of American English glides, Gick (2003) found that /w/ shows a similar timing pattern to /l/, with the labial gesture of /w/ (like the TT gesture of /l/) occurring earlier than the tongue dorsum gesture in syllable-initial position, and later in syllable-final position (where the labial gesture also displayed a reduction in magnitude). Given that both gestures have relatively wide constrictions (presumably [vocalic] in Sproat & Fujimura’s view), Gick (2003) proposed that the distinction between gestures must be more abstract (phonological) and language-specific, and assigns the category “C-gesture” to the lip gesture, and the category “V-gesture” to the tongue dorsum gesture for English /w/. The defining characteristics of a C-gesture in this account are “(1) final reduction, (2) intermediate magnitude under resyllabification, and (3) a tendency to occur farther away from the peak vowel.” (Gick, 2003, p.13). According to this proposal, any of the three gestures of /r/ could (theoretically) belong to either category so the expected timing relations are unclear; however, it is possible to make predictions about which category each gesture would belong to based on previous descriptions. First, as this account offers only two categories, the three gestures of /r/ should maximally display a two-way distinction in timing and magnitude patterns. Second, as to specific gestures, Delattre and Freeman (1968) and Zawadzki and Kuehn (1980) showed more lip rounding in prevocalic positions, suggesting that the labial gesture may act as a C-gesture; the more anterior lingual gesture is also likely to fall into the C-gesture category since (as noted above) Zawadzki and Kuehn (1980) observed a more advanced tongue position, Gick (1999) a greater gestural magnitude, and Delattre and Freeman (1968) an increased likelihood of a retroflex tongue shape, in syllable-initial position. If anything, then, Gick (2003) predicts that, for /r/, the lip and tongue blade will pattern together, both in showing final reduction and in occurring farther away from the peak vocalic element of the syllable.

In a study of timing patterns in liquids in six different languages, Gick et al. (2006) found that the tongue tip gesture in western Canadian English /l/ preceded the tongue dorsum gesture in prevocalic position, and followed it in postvocalic position, consistent with previous studies (though the prevocalic lag was greater, and the postvocalic lag smaller, than seen in studies of American English /l/). They concluded that perceptual recoverability (e.g., Chitoran, Goldstein, & Byrd, 2002; Kochetov, 2002; Mattingly, 1981; Silverman, 1997) plays a greater role in syllable-initial positions, while biomechanical factors such as the jaw cycle (e.g., Keating, 1983; Lindblom, 1983; MacNeilage, 1998) are more important in syllable-final positions. Specifically, according to Gick et al. (2006), perception-based studies predict that gestures in

Table 1

Summary of predictions of relative timing and magnitude by position

	Prevocalic	*Postvocalic*
Sproat & Fujimura (1993)	All three simultaneous All three reduced (if all three are [vocalic])	All three simultaneous TB reduced (if TB is [consonantal])
Browman & Goldstein (1995)	All three simultaneous	Dependent on relative magnitude of gestures
Gick (2003)	Lip & TB pattern together TR possibly reduced	Lip & TB pattern together Lip & TB reduced
Gick et al. (2006)	All three simultaneous	TR > TB > Lip

syllable-initial position should tend to be realized simultaneously. Biomechanically based analyses, on the other hand, in which the cyclical movement patterns of the jaw determine timing, predict that more anterior gestures will occur when the jaw is at its highest position (of either opening or closing), thus restricting anterior gestures to the time in the syllable furthest from the vocalic peak. Less anterior gestures, being located closer to the 'hinge' of the jaw, and moving in a direction essentially perpendicular to the primary jaw movement, should be less constrained by jaw position. Thus, according to Gick et al. (2006), we should expect to see all three gestures of /r/ occurring simultaneously in syllable-initial position, while in syllable-final position, the gestures should occur in back-to-front order (TR, then TB, then Lip).

In addition to the above studies, other studies have associated different gestures with different positions in a hierarchical model of the syllable (e.g., Carter, 1999, 2002). While the results of the present study may be found to bear on these issues, it is difficult to draw specific predictions regarding timing from such proposals.

Table 1 summarizes the predictions of each of the proposals described above with regards to /r/ in syllable-initial (prevocalic) and syllable-final (postvocalic) positions.

2 Method

An experiment was conducted using a combination of B/M-mode ultrasound imaging (for lingual data) and Optotrak tracking (for labial position data) to measure the movements of the three gestures of North American English /r/ in a variety of vocalic contexts at a natural speech rate.

2.1 Participants

Ten native speakers of Canadian English participated in this study, five female, five male, aged from 22 to 36. Of those who reported speaking other languages, three spoke French as a second language and one spoke Cantonese. Six were from Vancouver, two were from other parts of Western Canada, and two were from Ontario. Subjects were paid for their participation in this experiment. One of the male subjects from Vancouver was excluded from the analysis based on poor ultrasound image quality.

2.2 Stimuli

Stimuli were designed such that /r/ was flanked by maximally similar vowels. As illustrated below, the syllable position of the /r/ was varied such that it occurred in Initial (prevocalic) position, Final (postvocalic) position, and in a context where it could potentially be resyllabified (henceforth Resyllabifiable, postvocalic/word-final followed by a vowel-initial word). The Final condition included /h/, a segment with no oral gestures, following the target /r/, in order to prevent resyllabification. All vowels in Canadian English that normally occur word-finally (/i/, /e/, /a/, /o/, and /u/) were used as vocalic contexts, as each vowel has the potential to obscure one or more of the gestures of /r/.

Context for target /r/

The syllabic context for /r/ varied as follows:

a. Initial /r/	b. final (Resyllabifiable) /r/	c. Final /r/
...V_1#**R**V_1...	...V_1**R**#V_1...	...V_1**R**#hV_1...

Where V_1 = /i/, /e/, /a/, /o/, /u/

These stimuli were presented within the carrier phrase "... said x each ...," where x is an emphasized two-syllable nonsense phrase (with the /r/ in the middle) with equal stress on the two syllables. Recognizable monosyllabic words were used where possible in order to prompt the appropriate vowel and to ease the difficulty of the reading task. An item from a randomized list of names preceded each test phrase and an item from a randomized list of temporal nouns/noun phrases followed each test phrase. The presence of /r/ within the carrier sentences was avoided, particularly within syllables adjacent to the target phrase. Examples are given below:

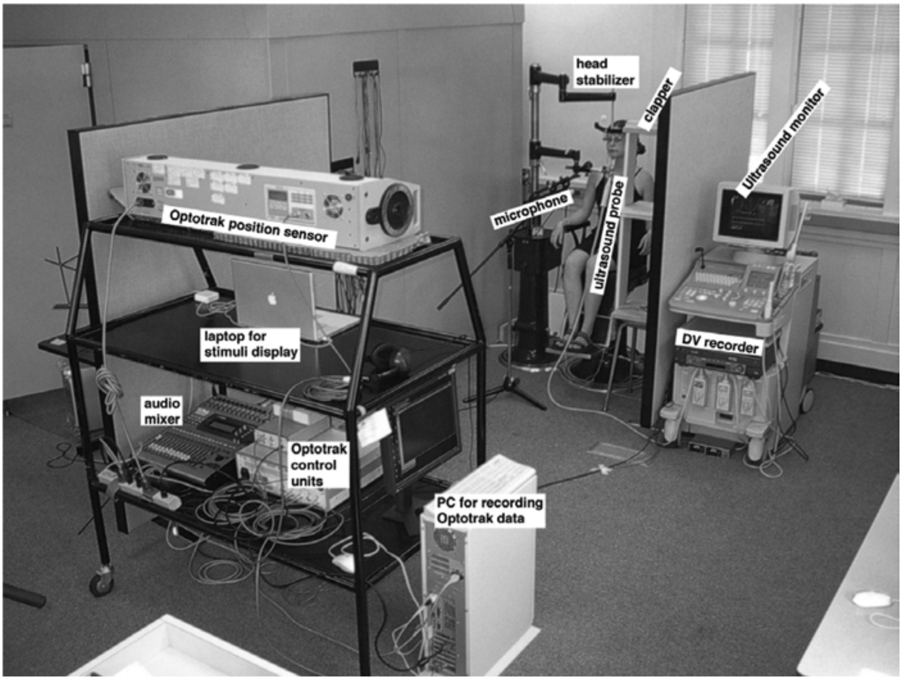
a. Initial /r/: Casey said "hay ray" each evening.
b. Resyllabifiable /r/: Mike said "hair A" each day.
c. Final /r/: Joan said "hair hay" each month.

The 150 sentences necessary for this experiment (10 tokens for each of 15 conditions [five vowel contexts, three syllable positions]) were randomized, along with 20 sentences required for a separate experiment that served as distractors. These sentences were divided into six sets and several additional distractor sentences were added to the beginning of each set in order to avoid list effects and to bring the number of sentences in each set up to 30. The full set of stimuli was therefore 180 sentences.

Not all of the data collected were included in the final analysis. After determining which vowel contexts were most suitable for observing and comparing all three gestures (i.e., TB, TR, Lip) across subjects, only the tokens in the vocalic contexts /e/ and /a/ were selected for analysis. For the nine speakers analyzed, movement associated with the TB gesture was visible in all three syllable contexts with the vowel /a/ and the TR gesture was similarly visible in the context of /e/. Further, the timing and magnitude of Lip movement associated with /r/ could be observed more easily with these vowels because they are unrounded in Canadian English. The relative timing of the two lingual gestures could then be compared using the labial gesture as a reference point.

Figure 1

Experimental set-up (from Wilson, 2006)



2.3 Procedure

Subjects were seated in a modified American Optical Co. ophthalmic examination chair (model 507-A) adjusted to maximize head stability and ultrasound probe contact. This included a two-point headrest located at the back of the head, just above the neck, and a two-point forehead stabilizing head restraint, which was secured in a position where it was in contact with the subject's head, but not with enough pressure to cause discomfort. The ultrasound transducer, mounted on a mechanical arm attached to the chair, was secured in a position where it pressed against the subject's neck in such a way as to provide a consistent midsagittal (B-mode) image of the subject's tongue from root to tip. Twelve infrared-emitting diodes (markers) were attached to the subject and apparatus. These were tracked by the three LED-sensing single-axis CCD cameras in the Optotrak camera bar, which was approximately 2 m in front of the subject at roughly head height. Subjects read from the 17-inch monitor of a laptop computer positioned below the Optotrak camera bar. Audio information was recorded via a microphone directly in front of the subject. Figure 1 (from Wilson, 2006) illustrates the experimental set-up.

Subjects were asked to read the stimuli sentences at a comfortable and natural rate as they were displayed on the monitor of the PowerBook G4 computer. A tone played for approximately 300 ms as each new sentence was presented. Sentences were displayed for three seconds each, with one second of a blank slide between each sentence. Presenting the stimuli in this manner, individually and just over 2 m from the subject at approximately eye level, has the effect of minimizing head movement (Stone, 2005). This is particularly important because no post-hoc correction was applied to the data based on head position. Work by Gick, Bird, and Wilson (2005) showed no evidence of a correlation between head position and tongue depth, as viewed with the ultrasound, and so no correction for submental tissue compression (from contact with the ultrasound transducer) was applied.

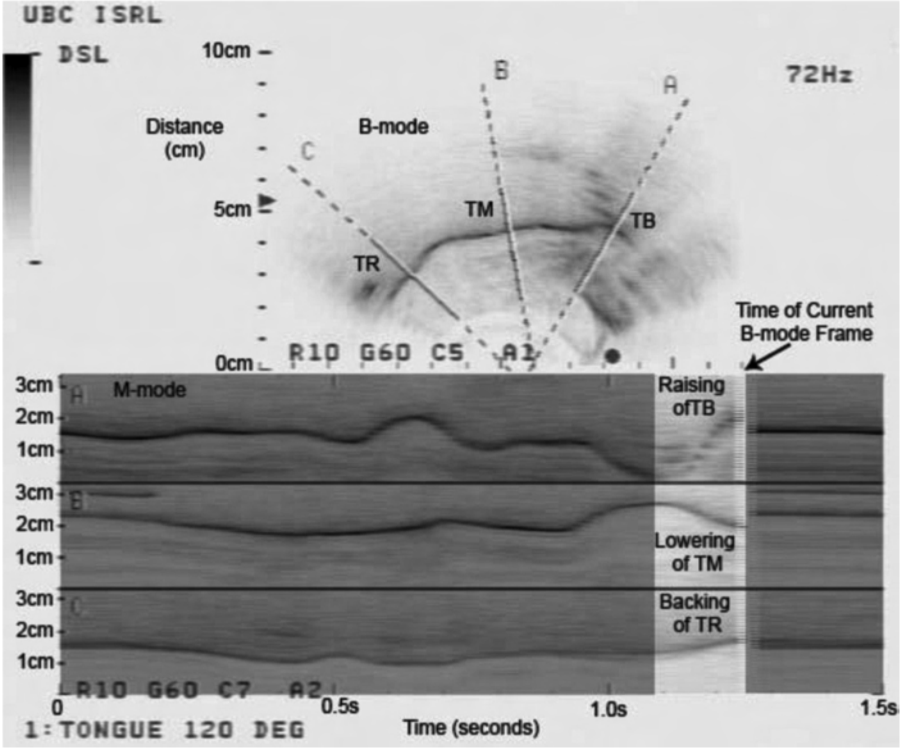
Prior to data collection, subjects were given a set of 15 practice sentences in order to familiarize them with the format of the sentences, get them settled into the chair, verify that the marker placements were secure, and allow time for ultrasound set-up adjustments. As described above, six sets of 30 sentences, lasting approximately 130 seconds/set were collected for each subject. Breaks between sets allowed for marker position to be verified and for data to be processed. Including set-up and breaks, the experiment took approximately 1.5 hours (per subject) to complete.

Ultrasound data were collected via an Aloka ProSound SSD-5000 ultrasound machine with a UST-9118 EV 180-degree probe/transducer. Ultrasound uses the echo patterns of ultra-high frequency sound both emitted and received by piezoelectric crystals contained in a small transducer. This signal is transmitted linearly through material of uniform density but reflects off air and is refracted when it meets bone. In (two-dimensional) B-mode, with the ultrasound transducer held under the chin and with the crystal array lying in the midsagittal plane of the head, the screen displays information about the superior surface of the tongue from the tongue root to near the tip (Stone, 1990) along the midsagittal plane. In combined B/M-mode, the B-mode midsagittal tongue line is displayed and its movements along one or more trajectories (chosen by the researcher) are tracked, smoothed, and presented visually as a continuous signal. Three cursors (A, B, C) were positioned such that they intersected with constrictions visible on the B-mode ultrasound image of the tongue in order to track the movements of the individual articulators shown in the M-mode signal (see Figure 2). Cursor A was placed so as to intersect the tongue blade/body (TB) between the tongue tip and the tongue mid, B was placed so as to intersect the tongue mid, located approximately in the uvular region of the tongue, and C was placed so as to intersect the tongue root (TR), often at a point as far back as was visible throughout the utterance. M-mode windows correspond approximately to the solid sections of lines A, B, and C.

Because the exact locations of the relevant lingual events for /r/ vary across subjects, cursor positions were determined based on constriction locations observed in the subjects' practice utterances on the monitor at the beginning of the experiment. Once fixed, cursor positions were constant throughout the rest of the data collection session. This method highlights a significant departure from lingual point tracking methods (EMMA, x-ray microbeam), in that tongue movement is measured at fixed constriction locations along the vocal tract rather than predetermined points on the surface of the tongue. Sweep speed (the rate at which the M-mode data is displayed/refreshed on the ultrasound monitor), was set at the highest setting (1.5 seconds per period) in order to have the most detailed data possible available in the exported video. The range (the total real distance represented in the window on the screen), for both the B-mode and M-mode displays, was set at 10 cm. The 10 cm of the M-mode display space was divided between the three cursors, meaning that each cursor could track movements over about 3.333 cm. These settings allowed for the whole tongue

Figure 2

Labeled B/M-mode ultrasound image of /r/ during the phrase *said "hoar owe"*

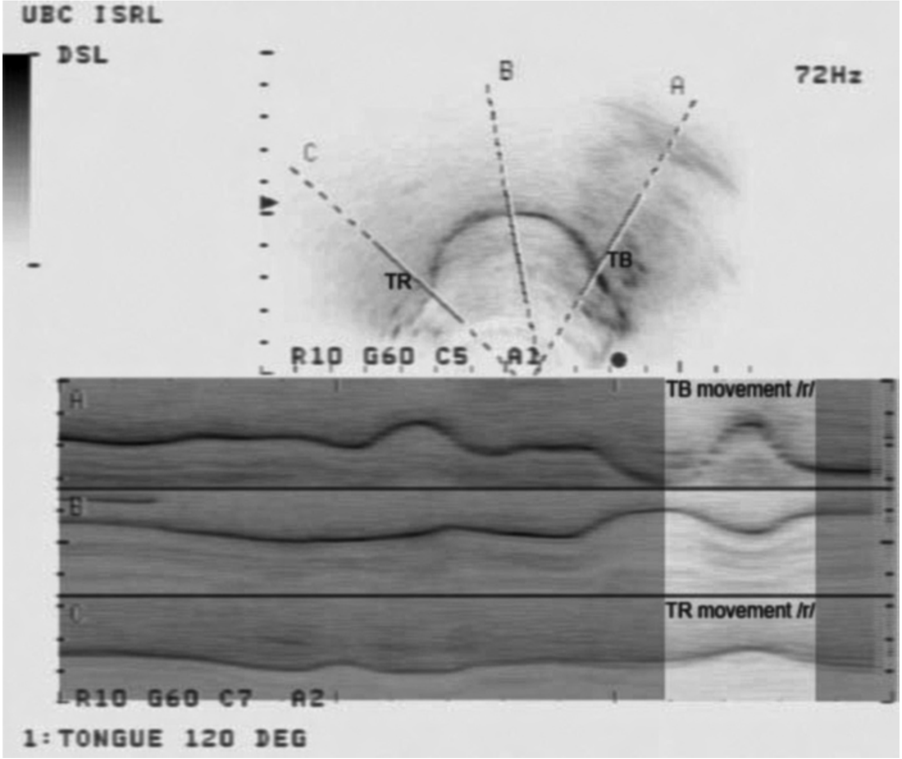


surface to be imaged in B-mode and the full range of motion to be tracked in M-mode. Resulting data were recorded to DV cassette along with a synchronized audio signal (via a Yamaha 01V digital mixing console) and subsequently loaded onto a Macintosh G4 computer using Adobe® Premiere® (version 6.0). Single frames clearly showing the complete M-mode traces of each instance of /r/ to be analyzed were exported as PICT files and analyzed in Adobe® Photoshop® (version 7.0.1). An example of a complete frame is given in Figure 3.

Three-dimensional positions of 12 infrared-emitting diodes attached to the lips, head, and apparatus were recorded using an Optotrak 3020 system (Northern Digital Inc.) in conjunction with Collect (version 2.002, Northern Digital) Optotrak software. For this experiment Optotrak data was collected at 90 Hz. As illustrated in Figure 4 (from Wilson, 2006) markers 1–4 were attached to a modified pair of glasses worn by the subject to track head position throughout the trials. Markers 5 and 6 were attached to the transducer (7 cm and 14 cm from the tip) to provide a stable line of reference in space. Marker 7 was mounted on a small piece of open cell foam and attached below the chin on the jawbone to provide information on jaw movement, and markers 8–11 were placed at the corners and the midline of the top and bottom lips. The final marker (12) was placed on a hinged wooden "clapper," which was used to synchronize the Optotrak and ultrasound signals.

Figure 3

Frame showing M-mode trace of /r/ during the phrase *said "hoar owe"*

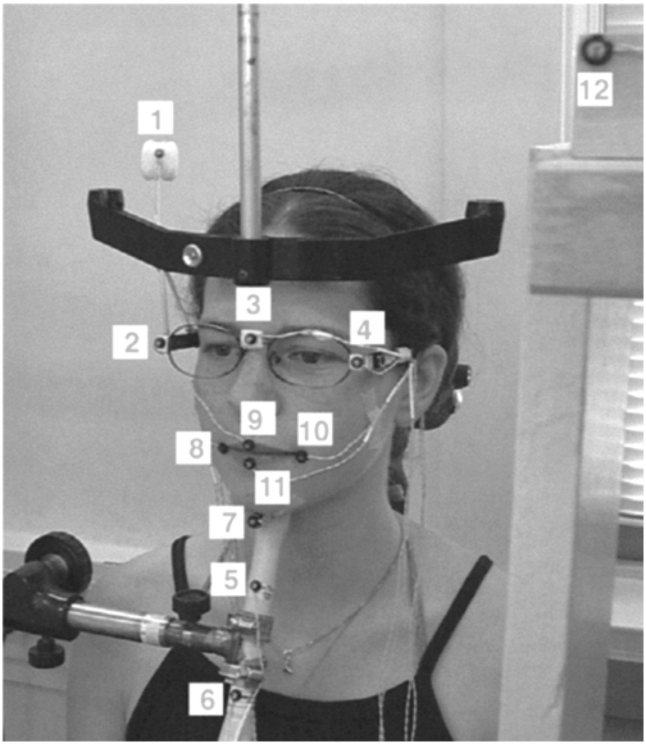


Marker position data (x, y, and z coordinates with greater than 0.1 mm accuracy) were recorded to a Micron Millennia XKU 333 computer. Data collected were monitored in real time for missing values (which could be due to an obstruction, a change in angle, and/or marker detachment) during each 130 trial. Proprietary software was used to convert raw camera sensor values to 3D positions (16-bit precision). During conversion, the position data were reoriented to a new centered coordinate frame. Converted data consisting of 90 x–y–z position values per second for each marker, were then exported for analysis in Microsoft® Excel® (version 10.1.0). Orientation of Optotrak data is x-vertical, y-horizontal, and z-depth.

A super-cardioid (Sennheiser 416) microphone was used to send the audio signal to a Yamaha 01V digital mixing console. In order to synchronize the signals, separate identical audio signals were then recorded with both Optotrak and ultrasound data signals via the mixer. The clapper (with Optotrak marker attached) was used at the beginning and end of every trial to set a 0 point and an end point for the synchronization of the Optotrak and audio signals. By comparing the time between the 0 point and the end point in the audio from the ultrasound recording, audio from the Optotrak, and the actual marker position in the Optotrak signal, it was possible to verify that no significant delay was introduced by the mixer.

Figure 4

Placement of Optotrak markers (from Wilson, 2006)



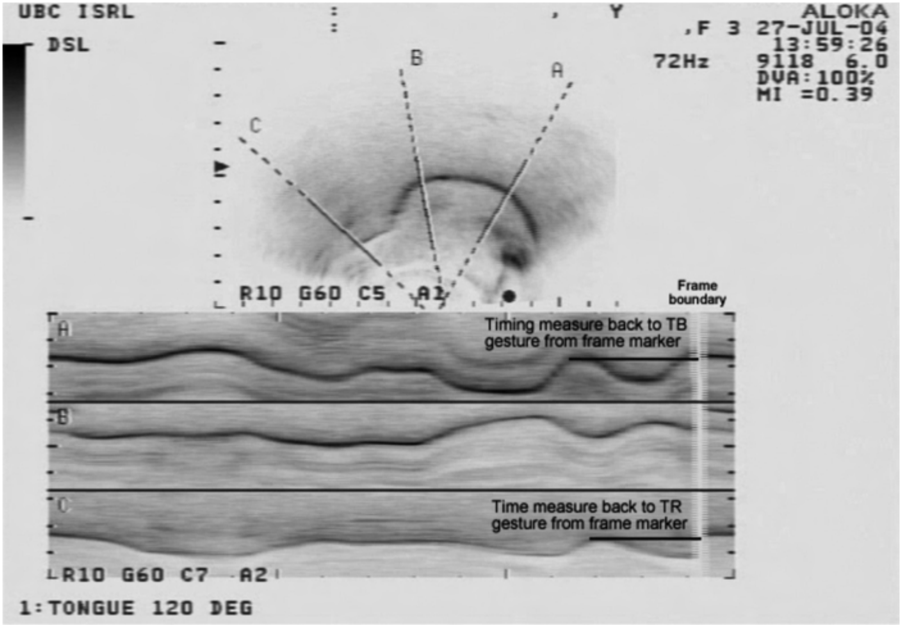
2.4 Analysis

PICT files of ultrasound video frames were opened in Adobe Photoshop, which had been set up to automatically scale the images based on the ultrasound's built-in timescale (visible along the M-mode window on the ultrasound display). The time in milliseconds from the frame marker back to the time when the identifiable movement associated with the /r/ gesture was completed was then measured using Photoshop's rectangular selection tool, as in Figure 5.

The gesture was considered to be "completed" at the time when it first reached a point within 5% of its peak constriction, based on the total range of movement of the articulator during speech. Based on the time of the /r/ gestures in the ultrasound signal relative to the 0 point described above, the point of closest approximation of the lips was located, and the time at which the lips came within 5% of their peak constriction (also based on the total range of movement during speech) was recorded as the time the gesture was completed. The times obtained for the Lingual gestures were then subtracted from the corresponding times for the Lip gesture, thus giving a measure of the difference in timing between the Lip and TB, and the Lip and TR. These differences could then be compared as relative timing offsets from the Lip gesture indicating time differences between the TB and TR.

Figure 5

Sample measures of timing: time is measured from frame marker back to A. TB gesture and B. TR gesture

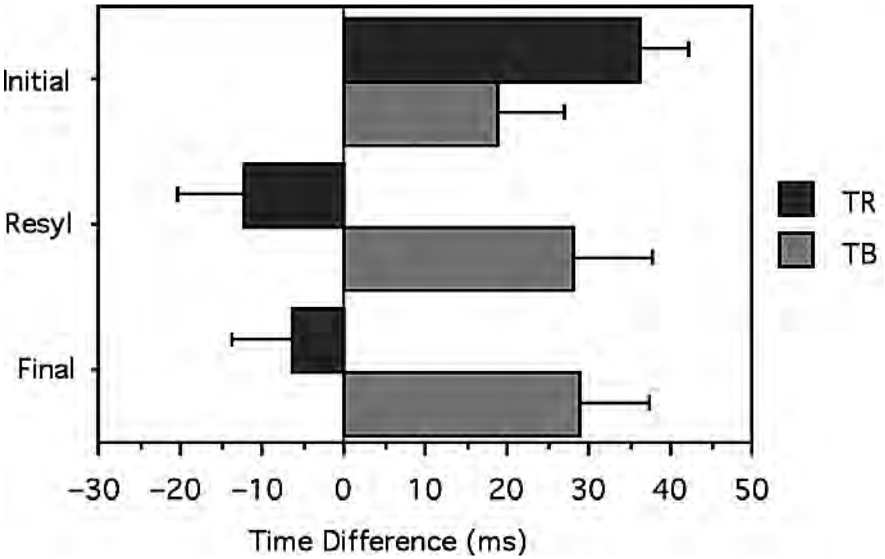


Magnitude was also measured for both Lingual and Labial gestures. Similar methods to those used for determining timing were employed to extract information about the relative size of the gestures across syllable positions. The same ultrasound frames that were used for the timing measures were used for magnitude measures, although magnitude measures were taken at the absolute peak of constriction. Photoshop was used to scale the images so that actual distances, based on the ultrasound display's scale, could be measured. Exactly the same scaling procedure was applied to all of the images for all of the subjects. The magnitude of gestures was measured as the distance from the lower border of the window (an arbitrary but consistently identifiable point) to the peak of the movement visible in the M-mode track for TB and TR. The relative magnitude across positions could then be compared. Magnitude of the Lip gesture was determined by finding the value for the extreme of approximation of the lips nearest the time of the /r/ (based on lingual timing data) in the Optotrak data. Degree of approximation was based on the Euclidean distance between the upper lip and lower lip markers (#9 & #11). This was calculated for the data using the vertical (x) and depth (z) dimension measures with the following equation:

$$d = \surd\,[\,(X_{UL} - X_{LL})^2 + (Z_{UL} - Z_{LL})^2\,]$$

Figure 6

Cross-subject timing of achievement of TB and TR gestures (relative to Lip), by position. Error bars indicate 95% confidence interval, 0 = time of Lip gesture



3 Results

3.1 Timing

Data for all nine subjects were combined and *t*-tests were calculated to test for differences in timing between the achievement of the TB and TR gestures relative to the Lip gesture in Initial (e.g., haw raw), Resyllabifiable (e.g., har awe), and Final positions (e.g., har haw). In addition, one group *t*-tests were calculated to test for significant differences between the Lip and Lingual gestures in each position (using a mean of 0 for Lip values). The overall order of gestures found was Lip > TB > TR (front to back) in Initial position, TR > Lip > TB in Resyllabifiable position, and TR / Lip > TB in Final position (see Figure 6, Tables 2 and 3). One group *t*-tests were used to test for differences between the timing of the TB and Lip and the TR and Lip because the timing of these gestures was calculated based on a mean of 0 ms for Lip. Unpaired *t*-tests were used to test for differences between the TB and TR because in this case there are two mean numbers (each of which is relative to Lip) that must be compared. Finally, ANOVAs were used to compare the differences in timing between positions for each articulator.

As shown in Figure 6 and Tables 2 and 3, in Initial position there was a significant difference across subjects in timing between the Lip and the TB, one group *t*-test, $p < .0001$, with the Lip gesture preceding the TB gesture by an average of 20.8 ms, as well as a significant difference between the Lips and TR, one group *t*-test, $p < .0001$, with

Table 2

One group t-tests for differences between lingual gestures and Lip (Lip = hypothesized mean of 0; significance level $p < .05$)

	Mean (ms)	*p value*
Initial Lip/TB	20.804	$p < .0001$
Initial Lip/TR	36.12	$p < .0001$
Resyl Lip/TB	28.287	$p < .0001$
Resyl Lip/TR	10.996	$p = .0043$
Final Lip/TB	27.601	$p < .0001$
Final Lip/TR	6.284	$p = .0911$

Table 3

Unpaired t-tests for differences between TB and TR by position (significance level $p < .05$)

	TB mean (ms)	*TR mean (ms)*	*Difference (ms)*	*p value*
Initial TB/TR	20.804	36.120	15.316	$p = .0016$
Resyl TB/TR:	28.287	–10.996	39.283	$p < .0001$
Final TB/TR	27.601	–6.284	33.886	$p < .0001$

the Lip preceding the TR by an average of 36.1 ms. Based on this, the TB preceded the TR by an average of 15.3 ms, unpaired t-test, $p = .0016$.

In the Resyllabifiable position, there was a significant difference in timing between the Lip and the TB, one group t-test, $p < .0001$, with the Lip preceding the TB by an average of 28.3 ms. TR was also significantly different from Lip, one group t-test, $p = .0043$, preceding it by an average of 11.0 ms. TR therefore preceded TB by an average of 39.3 ms, unpaired t- test, $p < .001$. In contrast with the Initial position order (Lip > TB > TR), the TR preceded the Lip and TB in Resyllabifiable position, where the order was TR > Lip > TB.

In the Final position there was a significant difference between the Lip and TB gestures, one group t-test, $p < .0001$, with the Lip preceding the TB by an average of 27.6 ms. The difference between the Lip and TR was not significant, one group t-test, $p = .0911$, in this position, but the difference between the TB and the TR was significant, unpaired t-test, $p < .0001$, with the TR preceding the TB by an average of 33.9 ms.

The timing of the TB gesture relative to the Lip gesture was relatively constant across positions: the Lip always preceded the TB. The TR gesture, however, followed the Lip in Initial position and showed the reverse order, TR then Lip, in Final and Resyllabifiable positions.

ANOVA analysis showed significant differences in time of achievement of the TR gesture relative to Lip across positions, $F(2, 290) = 54.614$, $p < .0001$. Fisher's PLSD post-hoc analysis revealed significant differences between the Initial and Final conditions, $p < .0001$, and between the Initial and Resyllabifiable conditions, $p < .0001$, for the TR. Despite the above finding that the TR was significantly different from Lip in Resyllabifiable position but not in Final position, no significant difference was found between the Final and Resyllabifiable contexts, $p = .3462$, for the TR.

An ANOVA of the differences between the times of achievement of the TB gesture relative to Lip in the three positions did not provide significant overall results, $F(2, 198) = 1.073$, $p = .3441$.

3.2 Magnitude

Non-normalized data for all nine subjects were combined and ANOVAs were calculated in order to determine if significant differences in the magnitude of gestures in the three syllable positions were present. For two subjects no Lip gesture was observable in the postvocalic conditions, so for these subjects only data for the Initial condition was included in the analysis. A significant difference was found between Final/Resyllabifiable and Initial positions for all three articulators; however, no significant differences between the Final and Resyllabifiable conditions were evident (see Figures 7–9).

In Initial position both the TB and Lip gestures had a greater magnitude than in Final or Resyllabifiable positions, while the TR gesture showed the opposite pattern and was reduced in Initial position, as compared to Final and Resyllabifiable positions.

As can be seen in Figure 7, overall ANOVA results indicate significant variances in the magnitude of the Lip gesture across positions, $F(2, 539) = 105.042$, $p < .0001$. Fisher's PLSD post-hoc analysis indicate that the degree of Lip approximation was significantly greater in the Initial condition than in the Final or the Resyllabifiable condition, $p < .0001$. No significant difference was observed between the Final and Resyllabifiable conditions, $p = .4329$.

Figure 7

Lip aperture across syllable positions. Error bars indicate 95% confidence interval. Note: For the Lip measures, 'aperture' (the distance between the lip markers) is represented, such that smaller values indicate greater constrictions

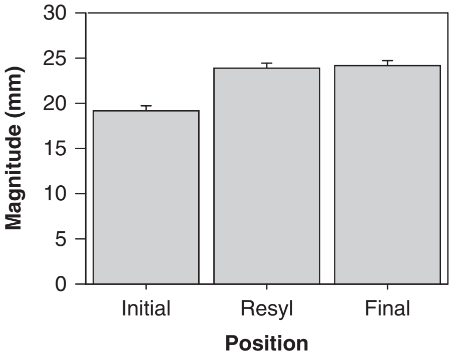
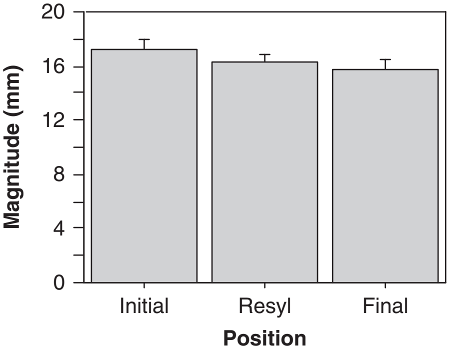


Figure 8

Magnitude of TB gesture across syllable positions. Error bars indicate 95% confidence interval



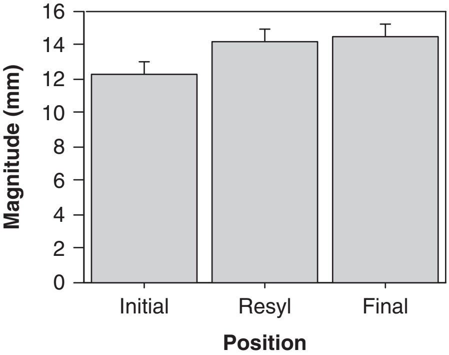
As can be seen in Figure 8, overall results indicate significant variances in the magnitude of the TB gesture across positions, $F(2, 276) = 4.963$, $p = .0076$. Fisher's PLSD post-hoc analysis indicates that the TB gesture was significantly greater in Initial position than in the Final, $p = .0020$, or Resyllabifiable, $p = .0461$, conditions.

The mean for the TB gesture was slightly higher for the Resyllabifiable condition, Mean = 16.3 mm, than the Final condition, Mean = 15.7 mm, but the difference was not significant, $p = .5978$.

Overall ANOVA results indicate significant variances in the magnitude of the TR gesture across positions, $F(2, 298) = 10.729$, $p < .0001$, as in Figure 9. Fisher's PLSD post-hoc analysis indicates that the magnitude of the TR gesture was significantly less in Initial position than it was in Final, $p < .0001$, or Resyllabifiable, $p = .0003$, conditions. There was no significant difference between the magnitudes for Final and Resyllabifiable conditions, $p = .5978$, which had a mean difference of only 0.3 mm.

Figure 9

Magnitude of TR gesture across syllable positions. Error bars indicate 95% confidence interval



3.3 Timing–magnitude interaction

As at least one hypothesis being tested in this article predicts a dependency between intergestural timing and gestural magnitude in at least one syllable position (Browman & Goldstein, 1995), a multivariate analysis of variance (MANOVA) was used to test for interactions between these dependent variables. Results suggest a significant correspondence between timing and magnitude across all syllable positions and articulators (Syllable Position: $F(2, 463) = 9.57$, $p < .0001$; Articulator: $F(1, 463) = 37.59$, $p < .0001$; Syllable Position × Articulator: $F(2, 463) = 29.63$, $p < .0001$). Since Browman and Goldstein (1995) only predict this interaction for syllable-final allophones, additional MANOVA tests were separately run on each syllable position. Results indicate a significant interaction between intergestural timing and gestural magnitude (across articulators) in all three syllable positions (Initial: $F(1, 160) = 8.24$, $p = .0047$; Resyllabifiable: $F(1, 145) = 44.86$, $p < .0001$; Final: $F(1, 158) = 36.83$, $p < .0001$).

4 Discussion

In this section, the results are further examined and the predictions shown in Table 1 are reviewed in light of these results for Canadian English /r/.

4.1 Summary of results

The syllable position-based differences observed in the overall results for Initial and Final positions were:

Initial position (e.g., a#ra):

Timing: strictly front-to-back (Lip > TB > TR)
Magnitude: Reduction of TR gesture.

Final position (e.g., ar#ha):

Timing: TR and Lip preceded TB (TR/Lip > TB)
Magnitude: Reduction of TB and Lip gestures.

MANOVA results further indicated a significant interaction between intergestural timing and gestural magnitude in all three syllable positions tested.

It is necessary to point out that the results for the Resyllabifiable (e.g., ar#a) condition are inconsistent, and neither clearly distinguishable from results for the Final condition, nor exactly the same. Most subjects inserted glottal stops between the final /r/ and the following vowel in the Resyllabifiable context at least some of the time (which explains the tendency of the /r/ to pattern as Final). However, the two conditions were not combined because there may have been categorical differences from the Final condition. The major difference between the two conditions was that there is a three-way distinction for Resyllabifiable and a two-way distinction for Final. This difference is the result of the TR being significantly different from Lip in the Resyllabifiable condition but not in the Final condition. It should be noted, however, that the difference between the TR across the two positions is not significant, and the amount of potential error in timing calculations between labial and lingual gestures exceeds the timing difference observed between the TR and the Lip in the Resyllabifiable context. Thus, it is not clear whether in fact the TR actually occurs

Table 4

(modified from Table 1)
Summary of predictions of relative timing and magnitude by position (with results)

	Prevocalic/initial	*Postvocalic/final*
Sproat & Fujimura (1993)	All three simultaneous All three reduced (if all three are [vocalic])	All three simultaneous TB reduced (if TB is [consonantal])
Browman & Goldstein (1995)	All three simultaneous	Dependent on relative magnitude of gestures
Gick (2003)	Lip & TB pattern together TR possibly reduced	Lip & TB pattern together Lip & TB reduced
Gick et al. (2006)	All three simultaneous	TR > TB > Lip
Present study	Lip > TB > TR TR reduced Significant interaction between timing and magnitude	TR / Lip > TB Lip & TB reduced Significant interaction between timing and magnitude

simultaneously with the Lip or before it. The remainder of this discussion will focus on the results for Initial and Final positions.

4.2 Comparison of results with predicted patterns

Several proposals that make specific predictions about the timing and magnitude of the gestures of /r/ were discussed above. Table 4 repeats the summarized predictions given in Table 1, with the addition of the cross-subject results from this study.

Generally speaking, the observed gestural reduction patterns are not unexpected, given that they match the predictions of at least one proposal (Gick, 2003), while the results for the relative timing of gestures seem to pose more interesting problems for previous analyses.

Both studies that state explicit hypotheses regarding reduction in Initial position, Sproat and Fujimura (1993) and Gick (2003), correctly predict the observed TR reduction in this position. Sproat and Fujimura (1993) also expect reduction in magnitude of the Lip and TB gestures (compared to final position) but this is not consistent with the present results. While Gick (2003) accurately predicts the reduction of the Lip and TB in Final position, the findings of the present study regarding timing patterns are not consistent with Gick's (2003) proposal.

The front-to-back timing observed in Initial position in the present study was not expected by any of the proposals. Gick (2003) does predict a timing offset in prevocalic (Initial) position, and that this will involve the TB preceding the TR; however, this proposal (where gestures belong to one of two phonological categories) cannot support the observed three-way distinction. That the simultaneity predicted by Sproat and Fujimura (1993), Browman and Goldstein (1995), and Gick et al. (2006) for gestures in Initial position was not found in the present study may be less problematic than it appears, as all of the predictions were based on observations of other languages or dialects (primarily American English). It is possible that Canadian English differs consistently in this respect from American English (note that Gick et al. [2006] found a timing offset in Initial position for Canadian English /l/).

The relative timing of the two lingual gestures in Final position (TR > TB) is predicted by Gick et al. (2006) and is not incompatible with Gick (2003). However, neither of these can account for the Lip gesture occurring with the TR rather than the TB as both proposals are based on the fixed property of degree of anteriority.

An alternate interpretation of Sproat and Fujimura (1993) classifying the TB gesture as [consonantal], as discussed above, also fails to capture the observed pattern in that, while it fits the observed timing in postvocalic position, prevocalic timing and the gestural reduction patterns are not predicted.

Browman and Goldstein's (1995) proposal is the only one that is consistent with the idea that the relative degree of constriction between gestures could change across positions. In the present study, a significant reduction in the magnitude of the Lip gesture was observed in Final position (the constriction at the Lip is on average more than 5 mm wider in Final position than in Initial position) while the TR gesture in this position is on average 2 mm greater than in Initial position. The TB gesture was an average of 2 mm smaller in Final position than in Initial, a small enough difference that the reduction of the Lip gesture could reverse these gestures in terms of width and leave the Lip gesture closer to the TR gesture. These results suggest that articulatory timing may be dependent on actual relative constriction width (which varies by position), consistent with Browman and Goldstein (1995). MANOVA results confirmed this possibility, indicating a significant interaction (across syllable position and articulator) between intergestural timing and gestural magnitude. Contrary to Browman and Goldstein (1995), however, MANOVA results within syllable position indicated that this interaction obtains not just in final position, but in all three syllable positions tested.

5 Conclusion

The goal of this study was to evaluate previous explanations of syllable-based allophonic variation in gestural timing and magnitude via a study of the three gestures of North American English /r/: Lip, tongue body (TB), and tongue root (TR). Overall, timing was observed to proceed sequentially from front-to-back in syllable-initial position, while in syllable-final position the TR and Lip gestures preceded the TB gesture. In terms of magnitude, the two more anterior gestures (TB and Lip) exhibited a relatively reduced magnitude in final position, while the least anterior gesture (TR) showed magnitude reduction in syllable-initial position. Further, a significant interaction was observed between intergestural timing and gestural magnitude. These findings taken together are not entirely consistent with any of the theories examined that attempted to explain syllable-based allophonic variation based on observation of two gestures. However, if it were extended to include syllable-initial position, Browman and Goldstein's (1995) proposal that constriction width predicts gestural timing patterns could account for the present results for Canadian English.

While this extension of Browman and Goldstein's position offers a promising account for the present data, testing of this proposal will require comparisons not just of gestural magnitude (as in the present study), but of the actual size of constrictions.

Future work of this kind may be able to take advantage of recent advancements in fast MRI technology.

A notable finding of this study is that the three-way distinction in timing between the Lip, TB, and TR gestures in Initial position cannot be represented in terms of a single binary phonological categorization of gestures.

References

ALWAN, A., NARAYANAN, S., & HAKER, K. (1997). Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part ii. The rhotics. *Journal of the Acoustical Society of America*, **101**(2), 1078–1089.

BROWMAN, C. P., & GOLDSTEIN, L. (1992). Articulatory phonology: An overview. *Phonetica*, **49**, 155–180.

BROWMAN, C. P., & GOLDSTEIN, L. (1995). Gestural syllable position effects in American English. In F. R. Bell-Berti & J. Lawrence (Eds.), *Producing speech: Contemporary issues. For Katherine Safford Harris*. Woodbury, NY: American Institute of Physics.

CARTER, P. (1999). Abstractness in phonology and extrinsic phonetic interpretation: The case of liquids in English. In J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A. C. Bailey (Eds.), *Proceedings of the XIVth International Congress of Phonetic Sciences* (pp.105–108). Berkeley: University of California Press.

CARTER, P. (2002). *Structured variation in British English liquids: The role of resonance.* Unpublished PhD dissertation, University of York, UK.

CHITORAN, I., GOLDSTEIN, L., & BYRD, D. (2002). Gestural overlap and recoverability: Articulatory evidence from Georgian. In C. Gussenhoven & N. Warner (Eds.), *Papers in laboratory phonology VII* (pp.419–447). Berlin/New York: Mouton de Gruyter.

DELATTRE, P., & FREEMAN, D. (1968). A dialect study of American r's by x-ray motion picture. *Linguistics*, **44**, 29–68.

DOCHERTY, G., & FOULKES, P. (2001). Variability in (r) production – instrumental perspectives. *Etudes & Travaux*, **4**(Dec), 173–184.

ESPY-WILSON, C. (2004). Articulatory strategies, speech acoustics and variability. In *Proceedings of From Sound to Sense, 50+ years of discoveries in speech communication*, Cambridge MA: MIT.

ESPY-WILSON, C. Y., BOYCE, S., JACKSON, M., NARAYANAN, S., & ALWAN, A. (2000). Acoustic modeling of American English /r/. *Journal of the Acoustical Society of America*, **108**, 343–356.

FOUGERON, C., & KEATING, P. (1997**)**. Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America*, **101**, 3728–3740.

GICK, B. (1999). A gesture-based account of intrusive consonants in English. *Phonology*, **16**, 29–54.

GICK, B. (2003). Articulatory correlates of ambisyllabicity in English glides and liquids. In J. Local, R. Ogden, & R. Temple (Eds.), *Papers in Laboratory Phonology VI: Constraints on phonetic interpretation* (pp.222–236). Cambridge: Cambridge University Press.

GICK, B., BIRD, S., & WILSON, I. (2005). Techniques for field application of lingual ultrasound imaging. *Clinical Linguistics and Phonetics*, **19**, 503–514.

GICK, B., CAMPBELL, F., OH, S., & TAMBURRI-WATT, L. (2006). Toward universals in the gestural organization of syllables: A cross-linguistic study of liquids. *Journal of Phonetics*, **34**, 49–72.

GUENTHER, F., ESPY-WILSON, C., BOYCE, S., MATTHIES, M., ZANDIPOUR, M., & PERKELL, J. (1999). Articulatory tradeoffs reduce acoustic variability during American English /r/ production. *Journal of the Acoustical Society of America*, **105**, 2854–2865.

HAGIWARA, R. (1995). Acoustic realizations of American /r/ as produced by women and men. *UCLA Working Papers in Phonetics*, **90**, 1–187.

HASHI, M., HONDA, K., & WESTBURY, J. (2003). Time-varying acoustic and articulatory characteristics of American English [r]: A cross-speaker study. *Journal of Phonetics*, **31**, 3–22.

KEATING, P. (1983). Comments on the jaw and syllable structure. *Journal of Phonetics*, **11**, 401–406.

KOCHETOV, A. (2002). *Production, perception, and emergent phonotactic patterns: A case of contrastive palatalization*. New York and London: Routledge.

KRAKOW, R. A. (1989). *The articulatory organization of syllables: A kinematic analysis of labial and velar gestures*. PhD dissertation, Yale University, USA.

KRAKOW, R. A. (1999). Physiological organization of syllables: A review. *Journal of Phonetics*, **27**, 23–54.

LINDAU, M. (1985). The story of /r/. In V. Fromkin (Ed.), *Phonetic linguistics: Essays in honor of Peter Ladefoged* (pp.157–168). London: Academic Press.

LINDBLOM, B. (1983). Economy of speech gestures. In P. F. MacNeilage (Ed.), *The production of speech* (pp. 217–246). New York: Springer-Verlag.

MacNEILAGE, P. F. (1998). The frame/content theory of evolution of speech production. *Brain and Behavioral Science*, **21**, 499–546.

MATTINGLY, I. G. (1981). Phonetic representation and speech synthesis by rule. In J. L. T. Myers & J. Anderson (Eds.), *The cognitive representation of speech* (pp.415–420). Amsterdam: North Holland.

SILVERMAN, D. (1997). *Phasing and recoverability*. Outstanding dissertations in Linguistics. New York: Garland.

SPROAT, R., & FUJIMURA, O. (1993). Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics*, **21**, 291–311.

STONE, M. (1990). A three-dimensional model of tongue movement based on ultrasound and x-ray microbeam data. *Journal of the Acoustical Society of America*, **87**, 2207–2217.

STONE, M. (2005). A guide to analysing tongue motion from ultrasound images. *Clinical Linguistics and Phonetics*, **19**, 455–502.

TIEDE, M., BOYCE, S., HOLLAND, C., & CHOE, A. (2004). A new taxonomy of American English /r/ using MRI and ultrasound. Poster presented at the *147th Meeting of the Acoustical Society of America*. Retrieved 2 June, from http://scitation.aip.org/confst/ASA/data/1/5pSC37.pdf

ULDALL, E. (1958). American ‘molar’ R and ‘flapped’ T. *Revista do Laboratorio de Fonetica Experimental da Faculdade de Letras Da Universidad de Coimbra*, **4**, 10310–6.

WESTBURY, J. R., HASHI, M., & LINDSTROM, M. J. (1998). Differences among speakers in articulation of American English /r/. *Speech Communication*, **26**, 203–226.

WILSON, I. L. (2006). *Articulatory settings of French and English monolingual and bilingual speakers.* Unpublished PhD dissertation, University of British Columbia, Canada.

ZAWADZKI, P. A., & KUEHN, D. P. (1980). A cineradiographic study of static and dynamic aspects of American English /r/. *Phonetica*, **37**, 253–266.