

# THREE DIMENSIONAL ULTRASOUND IMAGING OF PRE- AND POST-VOCALIC LIQUID CONSONANTS IN AMERICAN ENGLISH: PRELIMINARY OBSERVATIONS

*Kelly H. Berkson, Kenneth de Jong*

Indiana University  
Department of Linguistics  
Bloomington, Indiana

*Steven M. Lulich*

Indiana University  
Department of Speech and Hearing Sciences  
Bloomington, Indiana

## ABSTRACT

This study presents articulation data of American English laterals and rhotics captured by combining real-time 3D ultrasound, digitized 3D palate impressions, and time-aligned audio recordings. While articulation of laterals and rhotics has long been of interest, traditional two-dimensional imaging techniques are subject to limitations because the vocal tract is three-dimensional. The technological advances rendering 3D imaging possible therefore enable more sophisticated understanding of these complex articulations. Results from two speakers are presented, one from New England and one from Indiana. Lateral productions are different for each speaker, but both speakers exhibit mirrored temporal phasing of coronal and dorsal occlusions in /l/ in onset vs. coda position, with dorsal articulations occurring temporally closer to the neighboring vowel. Lingual shape for rhotic productions is complex but consistent across speaker, syllable position, and vowel context.

**Index Terms**— ultrasound, vocal tract imaging, laterals, rhotics

## 1. INTRODUCTION

Typical introductions to articulatory phonetics present a basic representational scheme for describing vocal tract shapes for consonants, generally that of the mid-sagittal diagram. Implicit in this representation is the idea that, with the notable exception of lateral consonants, conceptions of speech articulation can be simplified to configurations of the vocal tract mid-line. This simplified two-dimensional representation is not without benefits. Many very successful simulations of articulatory-to-acoustic relationships over the years have also tended to simplify the basic effects of constrictions at various points along the vocal tract as if occurring within cylindrical tubes [1, 2], essentially mirroring the idea that a cross-sectional distance is sufficient to describe the articulatory effects on the vocal tract resonators [3].

The vocal tract is not two-dimensional, however, and recent technological advancements allow us to step away

from 2D representations of speech articulation and to begin to examine speech production as a three-dimensional phenomenon. Tongue shapes viewed from this perspective suggest very different interpretations of well-known phonetic behaviors.

The targets of the current study are the so-called ‘liquid’ consonants in American English (AE), referring here to the non-glide approximants and serving as a cover term for laterals and rhotics. The class of laterals and rhotics is notorious for comprising a broad array of phonologically difficult and articulatorily unusual consonants, which have traditionally been described in AE as involving multiple articulations - in contrast with other consonants. The AE lateral is noted as having a ‘dark’ allophone in post-vocalic position, where ‘dark’ is a term referring to the acoustic effect of a secondary articulation in the uvular or velar region, which causes extreme and selective lowering of the second formant. Some AE varieties have generalized this production to pre-vocalic positions as well [4], yielding the general description of laterals in AE as being composed of two gestures, one coronal and one dorsal [5]. Similarly, early X-ray studies of AE rhotics have indicated a very common articulatory pattern involving coordinated constrictions at the palate and in the mid-to-lower pharynx [6, 7].

Both laterals and rhotics are therefore thought to involve double articulations. Two aspects of liquid sound production are of particular interest in this respect. The first is coordination. Sproat and Fujimura’s classic work on the articulation of laterals discusses in detail the question of how multiple constrictions are coordinated in the case of AE /l/ [5]. As part of a much broader discussion of the large problem of understanding the sequencing, temporal binding, and coordination of speech articulations, multiple constrictions in a single segment present an excellent test case for examining how disparate constriction gestures can be bound together into a functional unit. Sproat and Fujimura’s original observations on /l/ indicate asynchronous timing of the two gestures, such that the coronal articulation precedes the dorsal articulation when /l/ is pre-vocalic while the coronal articulation follows the dorsal articulation when the /l/ is post-vocalic. What

this systematic pattern suggests is that intra-segmental coordination reflects larger temporal coordinative influences which produce segments within larger syllabic composites.

No such observations have been made concerning the gestural components of the rhotic, but a second fact about AE liquids is that they exhibit a striking pattern of variation. The rhotic is most well-known in this regard, on the basis of Delattre and Freeman's original 1968 mid-sagittal X-ray study [6]. This work, which includes imaging of American English /r/, noted a striking difference in mid-sagittal posture between speakers: some speakers assumed a retroflexion of the tongue tip while others exhibited a combination of palatal and pharyngeal lingual constrictions, often accompanied by lip rounding. Other speakers used both postures, with the retroflex appearing pre-vocally and the complex dorsal articulation appearing post-vocally. The typical analysis of this behavior, explored in great detail by [8], is that the two articulations yield a very similar acoustic pattern, and hence the variation in articulation is held up as a prime example of acoustically-based motor equivalence. That is, the facts about the rhotic support the contention that the motor tasks around which articulatory behavior are constructed are acoustic in nature, rather than just articulatory, as is modeled in such approaches as Task Dynamics [9] and Articulatory Phonology [10].

In sum, then, liquids involve multiple constrictions and apparently acoustically-oriented variability. The challenge, thus far, has been that the sophistication with which we can investigate the acoustic correlates of these articulatory settings has outpaced our capacity to render finely-detailed three-dimensional articulatory images. Three-dimensional imaging can dramatically enhance our investigation of multiple-constriction consonants, temporal coordination within them, and variation across different vocalic contexts. 3D MRI imaging, for instance, has improved our understanding of vocal tract shape and volume in the production of laterals [11] and rhotics [12]. Reconstructed 3D ultrasound images of 18 AE sounds, including the lateral and the rhotic vowel, have been similarly informative [13, 14], supporting the claim that the vocal tract reconstruction made possible by 3D imaging is enlightening. The technological advances which have occurred in recent years, meanwhile, render it possible to create imaging with greater temporal and spatial detail. With this in mind, the current project has begun collecting recordings of AE speakers producing laterals and rhotics in pre- and post-vocalic positions across varied vowel contexts. Findings reported here are for two pilot speakers.

## 2. METHODS

Two female speakers of American English participated in this study. One (author K.H.B.) is a native of New Hampshire (referred to as NH) and one is a native of southern Indiana (referred to as IN), but both speakers have residential histories

including multiple regions of the U.S.

The participants produced the following words in isolation: *keel, kill, call, leek, lick, lock, reek, rick, rock, roll, lore, coke*. Each word was separated by the phrase 'tea cod', which required the speakers to move all of the lingual articulators without producing liquid consonants.

Ultrasound images were recorded with a Philips EpiQ 7G system using an xMatrix x6-1 digital 3D/4D transducer. The transducer was secured under the chin using an Articulate Instruments ultrasound stabilization headset. Recording frame rates were approximately 10 volumes per second.

Subjects were seated in a sound-treated booth, and then cued with each word on a computer monitor. A SHURE KSM32 microphone was placed in front of the speakers on a stool. Audio and ultrasound recordings were synchronized by means of a foot pedal with USB interface to the ultrasound system and BNC interface to a LabView-based digital signal acquisition system running on a Windows 7 computer.

Uncompressed DICOM ultrasound files were transferred to the Windows 7 computer and exported to the binary FLD file format using Philips QLab software.

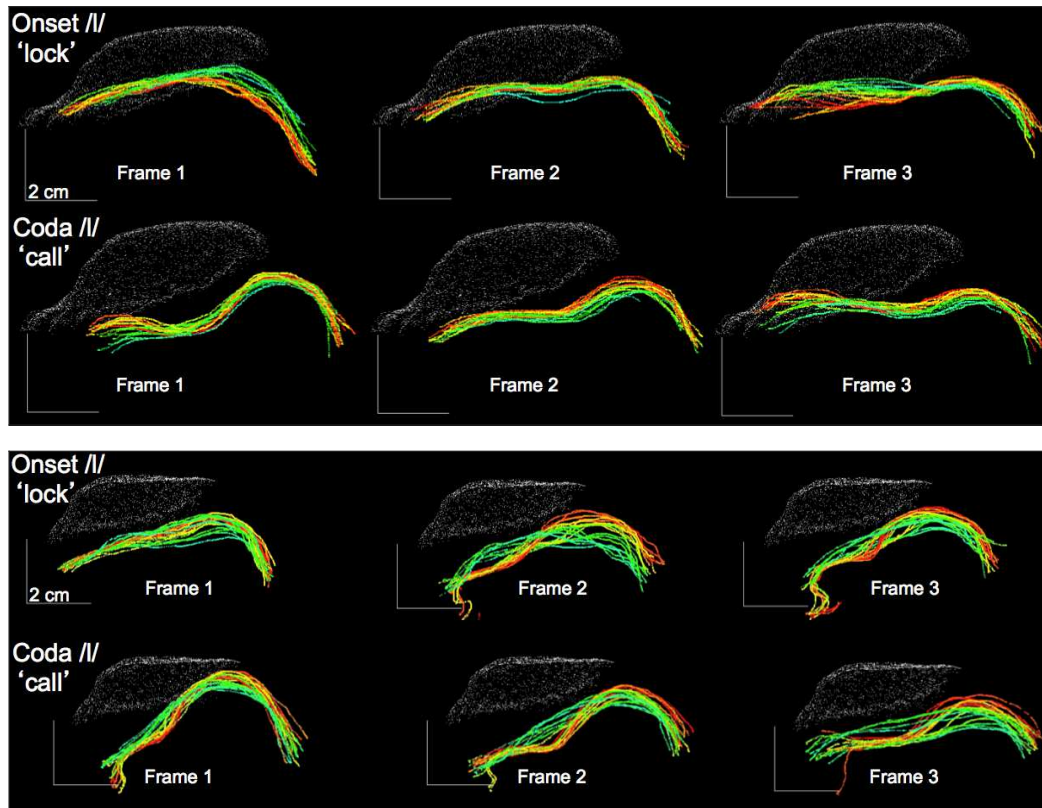
To help situate the ultrasound images in the context of the speakers' vocal tracts, palate impressions for each speaker were made using dental grade alginate and were digitized into STL formatted files with a NextEngine 3D laser scanner. The FLD ultrasound files and the STL palate files were analyzed using a custom MATLAB toolbox called 'WASL'.

Within WASL, palates were manually and subjectively registered with the tongue data, using the tongue-on-palate articulation of plosives in the 'teacod' frames as visual references. Sagittal ultrasound slices with superimposed palate outlines were converted to JPG images, uploaded to a web-accessed SQL database, and traced in the browser window by clicking-and-dragging with a mouse. Multiple sagittal slices were traced for each frame of each word, working with images as far off the mid-line as possible to allow visualization of the tongue surface in three dimensions.

## 3. RESULTS

### 3.1. Laterals

Images for three frames in NH's productions of /l/ in a pre-vocalic context (in 'lock') are presented in the top frames of Figure 1. Images show traces parallel to the sagittal plane, with red lines indicating traces in the mid-line, and green lines indicating increasingly lateral traces. The aligned palate is shown as a superimposed point cloud. The left-most frame shows the tongue at rest, while the right-most frame shows the full /l/ articulation. Pre-vocalic /l/ in NH's corpus conforms with the configuration found in the rightmost frame, where there are two parts of the tongue mid-line which extend upwards toward the opposite wall of the vocal tract. To the left is the anterior coronal contact with the alveolar ridge,



**Fig. 1.** Top: Three consecutive frames in NH's production of /l/ in 'lock' (upper frames) and 'call' (lower frames). Red traces indicate tongue mid-line, green traces lateral tongue margins. Bottom: Corresponding data for speaker IN.

apparent in the convergence of the red tongue trace and the palate. The second constriction is to the right, where the back of the tongue is domed (with the red trace more extreme than the green-traced margins of the tongue) toward the region of the upper pharynx. This dual-articulation configuration indicates that the pre-vocal /l/ is uvularized. This configuration is consistent across the corpus and is accompanied by a noticeable lowering of the second formant going into the vowel.

The second row of traces shows a sequence for NH's post-vocalic /l/, where the uvular constriction is much more pronounced. This is especially clear in the left-most frame, as is the cupping of the front section of the tongue. One other aspect of the pre- vs. post-vocalic distinction is very obvious here as well, and that concerns the relative timing of the anterior and posterior constrictions. In the pre-vocalic context, the two constrictions are roughly simultaneous with the coronal constriction slightly preceding the uvular constriction, while in the post-vocalic production the coronal constriction clearly lags behind the uvular constriction, by an amount on the order of 0.2 seconds.

Figure 1 also presents the parallel images for IN, shown in the bottom frames. Numerous differences between the speakers are apparent, including palate shape: IN shows a much flatter and shorter palate than NH, and a much less pronounced alveolar ridge. Also notable is that the IN lateral tongue shapes are quite similar in both pre- and post-vocalic position, although with opposite phasing. The general con-

figuration (frames 2 and 3 of 'lock' and frames 1 and 2 of 'call') consists of a dip in the mid-line of the anterior portion of the tongue, with the tip and dorsum raised somewhat from this position. This dip does not extend to the margins of the tongue, apparently representing a 'dimple' in the center of the tongue.

This general configuration was found in both speakers for both pre- and post-vocalic productions of the lateral. Also quite similar across the two speakers are the prominence of the uvular constriction in the post-vocalic position, with dorsal arching much more prominent in post-vocalic position. The tongue shapes are also very similar in both speakers, indicating similar temporal sequencing effects.

A notable difference between IN and NH concerns the anterior constriction: there is no clear evidence for anterior contact in IN's laterals in either pre- or post-vocalic position. This surprising finding was pervasive throughout IN's laterals, but without any discernable perceptual or acoustic abnormalities, and points to the suggestion that AE /l/ (even so-called 'light /l/') is darker than /l/ in many other languages [15]. To determine whether these images resulted from erroneous tongue-palate registration, a close examination of IN's productions of /d/ in teacod was conducted. This showed clear contact between the anterior portion of the tongue and the palate, indicating very good alignment of the lingual and palate data. Hence, if IN were making contact with the anterior portion of the tongue in the laterals, it would be visible here. Our

conclusion for IN is therefore that the lateral /l/ is pervasively articulated as a uvular approximant. At present, we are unsure how the acoustics are produced with this configuration, since the productions do not sound at all unusual or different from the lateral articulations produced by the other speaker. Particularly intriguing in this respect is that the cupping of the forward portion of the tongue is similar in both speakers.

### 3.2. Rhotics

Unlike laterals, lingual postures during rhotic productions were very similar in pre- and post-vocalic contexts. The overall posture for NH and IN are also similar, involving even more extreme cupping of the midline of the tongue, though with a generally more posterior location for the cup situated somewhere in the uvular region.

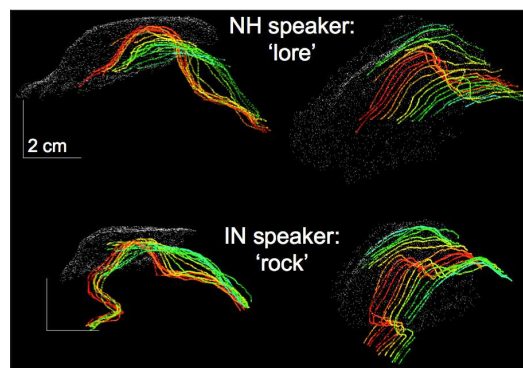
Examples of rhotic configurations from each talker are given in Figure 2, which shows a profile view of the tongue for one production of a post-vocalic /r/ for NH and of a pre-vocalic /r/ for IN. These productions are representative of the entire corpus. As before, midline traces are red and lateral traces are green. Rotated views of the same images are also shown.

The midline traces for both talkers consistently show the two-constriction profile found in the early X-ray studies [6], with both a clear palatal constriction and a clear curvature of the tongue root toward the back of the pharynx. However, as with the laterals, the profile of the tongue on the left and right margins differs dramatically from the profile shown at the midline. On the margins, there is a single curve with an apex at the location of the dip in the midline trace. The rotated images help make sense of this configuration as indicating a large dimple in the middle back of the tongue. Thus, what looks like two constrictions in the midline corresponds to a hollow in the center of the tongue body, when the tongue is viewed as the three-dimensional structure that it is. Viewed in this way, we can see that the pharyngeal constriction on the mid-line is emphasized by cupping of the dorsum

## 4. CONCLUSION

These findings indicate a number of promising lines of inquiry, which are currently being pursued with a larger and more systematic corpus. In addition to interesting questions concerning the relative timing and strength of the various constriction components and how they interact with the position of the consonant with respect to the vowel, a larger set of questions are raised by the three-dimensional aspect of the imaging.

One may note in the dynamic frames (Fig. 1) a diverging movement of the midline of the tongue relative to the margins of the tongue. Particularly intriguing are relations such that a lower midline in the front of the tongue corresponds to a higher midline in the posterior region (relative to the tongue



**Fig. 2.** Tongue shape in production of rhotics by NH (top) producing ‘lore’ and IN (bottom) producing ‘rock’. Images to the left present traces parallel to the sagittal plane. Mid-line traces are red, lateral traces are green. Palates are presented as point clouds. Images to the right are rotated with the front downward and to the left.

margins). This relation suggests that cupping may be indicating a compression of the forward part of the tongue which is executed in part in order to push the posterior region of the tongue upwards and backwards, since the tongue is a volume-preserving organ. Similarly, in the rhotic articulation (Fig. 2) the posterior dimple in the midline corresponds to an extension of the midline of the palatal region of the tongue. It is quite possible, then, that what in the older sagittal X-ray images appeared to be two constrictions may be better interpreted as one ‘un-constriction’, or compression of the midline of the tongue, which is part of a mechanism for extending the midline constriction of the tongue elsewhere. In any event, three dimensional imaging of the overall tongue shape pushes to the fore the role of the tongue as an articulator with intrinsic three-dimensional dynamics, and hence also the three-dimensional role of articulation in phonetics.

## 5. REFERENCES

- [1] K. N. Stevens, *Acoustic Phonetics*, MIT Press, Cambridge, MA, 1998.
- [2] G. Fant, *Acoustic Theory of Speech Production, Revised Edition*, Mouton de Gruyter, 1970.
- [3] R. S. McGowan, M. TT. Jackson, and M. A. Berger, “Analyses of vocal tract cross-distance to area mapping: An investigation of a set of vowel images,” *Journal of the Acoustical Society of America*, vol. 131.1, pp. 424–434, 2012.
- [4] J. Wells, *Accents of English*, Cambridge University Press, New York, NY, 1982.
- [5] R. Sproat and O. Fujimura, “Allophonic variation in English /l/ and its implications for phonetic implemen-

- tation,” *Journal of Phonetics*, vol. 21.3, pp. 291–331, 1993.
- [6] P. Delattre and D. Freeman, “A dialect study of American r’s by x-ray motion picture,” *Linguistics*, vol. 44, pp. 29–68, 1968.
  - [7] J. Westbury, M. Hashi, and M. Lindstrom, “Differences among speakers in lingual articulation for American English /r/,” *Speech Communication*, vol. 26, pp. 203–226, 1998.
  - [8] F. H. Guenther, C. Y. Espy-Wilson, S. E. Boyce, M. L. Matthies, M. Zandipour, and J. S. Perkell, “Articulatory tradeoffs reduce acoustic variability during American English /r/ productions,” *The Journal of the Acoustical Society of America*, vol. 105.5, pp. 2854–2865, 1999.
  - [9] E. Saltzman and K. G. Munhall, “A dynamical approach to gestural patterning in speech production,” *Ecological Psychology*, vol. 1, pp. 333–392, 1989.
  - [10] C. P. Browman and L. Goldstein, “Tiers in articulatory phonology, with some implications for casual speech,” in *Laboratory Phonology I*, J. Kingston and M. E. Beckman, Eds., pp. 341–376. Cambridge University Press, Cambridge, 1990.
  - [11] S. Narayanan, A. Alwan, and K. Haker, “Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part I. The laterals,” *Journal of the Acoustical Society of America*, vol. 101.2, pp. 1064–1077, 1997.
  - [12] A. Alwan, S. Narayanan, and K. Haker, “Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part II. The rhotics,” *Journal of the Acoustical Society of America*, vol. 101.2, pp. 1078–1089, 1997.
  - [13] M. Stone and A. Lundberg, “Three-dimensional tongue surface shapes of English consonants and vowels,” *Journal of the Acoustical Society of America*, vol. 99.6, pp. 3728–3737, 1996.
  - [14] A. Lundberg and M. Stone, “Three-dimensional tongue surface reconstruction: Practical considerations for ultrasound data,” *Journal of the Acoustical Society of America*, vol. 106.5, pp. 2858–2867, 1999.
  - [15] D. Recasens, “Darkness in [l] as a scalar phonetic property: implications for phonology and articulatory control,” *Clinical Linguistics and Phonetics*, vol. 18, pp. 593–603, 2004.