# An MRI study of articulatory settings of L1 and L2 speakers of American English

*Vikram Ramanarayanan, Louis Goldstein, Dani Byrd and Shrikanth Narayanan*
University of Southern California, Los Angeles, CA – 90007

## ABSTRACT

In this paper, we analyze the "articulatory setting" or "basis of articulation" of both L1 and L2 speakers of American English and investigate possible articulation parameters for its control within the framework of Task Dynamics [1]. More specifically, we extend methods and results presented in [2] to the case of native German, Hindi and Tamil speakers producing L2 English. We use audio-synchronized real-time magnetic resonance imaging (MRI) technology [3] to record all speakers while they produced sentences or paragraphs in the MRI scanner. We further describe the automatic extraction of measures to capture the vocal tract posture during unfilled pause intervals as well as at absolute rest. While the results indicate that individual speakers of these different native languages have significantly different articulatory settings when speaking English (which cannot be assumed to be due to language differences per se), all speakers exhibit two important effects. (1) The postures assumed during inter-speech pauses are significantly different than those assumed during an absolute rest position. (2) Based on analysis of variability during inter-speech pauses, it is more likely that individual *articulator* rest positions may be controlled to obtain these settings, rather than *constriction* (e.g., tract variable) positions.

With regard to speech planning and execution, it would be useful to understand whether articulatory setting (AS) is a phonological phenomenon or a functional target/by-product of the execution of the speech plan. Studies in the literature such as [4] have argued for the existence of a language-specific AS and have further speculated that speech rest positions are specified in a manner similar to actual speech targets. They found significant differences between speech rest positions assumed by English and French subjects (with measurements taken from x-ray data at 5 positions in the vocal tract). They further compared the standard deviations of vocal tract measurements taken during inter-utterance rest positions to those taken from the target vowel /i/ to test whether the accuracy of movements into an inter-utterance rest position was similar to that of a specified articulatory target and not just a transition point solely determined by the immediately surrounding sounds. They found no significant differences in the standard deviations of the two groups, leading them to suggest that a language's inter-speech posture may be specified as part of the phonetic or phonological inventory of the language in question. More recently, we proposed an automatic method using real-time MRI [3] to evaluate postural differences ("articulatory setting") between pauses in-between speech as opposed to an absolute rest position in both read and spontaneous speaking styles [2] and showed significant postural differences and differences in degree of active control between these different cases.

We used a database of read speech (TIMIT sentences, rainbow passage and north wind and the sun passage) elicited in turn from 5 native American English and L2 English speakers whose L1 was German(3), Hindi(1) and Tamil(1). The first four L1 English speakers spoke TIMIT sentences, while the rest spoke the passages; also note that all L1 English speakers were female, and all L2 speakers were male). We acquired audio-synchronized midsagittal real-time MR images of the vocal tract [3] with a repetition time of TR=6.5ms on a GE Signa 1.5T scanner with a 13 interleaf spiral gradient echo pulse sequence. The slice thickness was approximately 3mm. The sliding window reconstruction was at a rate of 22.4 frames per second. Field-of-view (FOV), which can be thought of as a zoom factor, was set depending on the subjects head size.

For each speaker's read speech data, we first extract pauses from these utterances and further extract measures of posture during these intervals as well as during absolute rest positions (which we compute from the first and last images of each sequence, where the speaker is not engaged in any speech activity). In order to obtain constriction-related postural measures, we divide the vocal tract into regions bounded by vocal tract constricting devices (such as the lips, tongue tip, tongue dor-

| Speaker | A1(I) | A1(R) | A2(I) | A2(R) | A3(I) | A3(R) | JA(I) | JA(R) |
|---|---|---|---|---|---|---|---|---|
| **Eng1(m1)** | -0.88 | -0.42 | -0.41 | -0.10 | 1.16 | -1.08 | 0.86 | -2.93 |
| **Eng2(s2)** | 0.32 | 3.74 | -0.51 | -0.79 | 0.52 | -1.24 | -1.23 | -2.55 |
| **Eng3(l4)** | 1.51 | 4.51 | 1.85 | 0.09 | 0.98 | -1.41 | 1.22 | -0.06 |
| **Eng4(h5)** | 1.65 | -0.55 | 2.03 | 0.96 | 1.96 | -2.14 | -0.60 | -0.43 |
| **Eng5(d6)** | -0.15 | -3.29 | -1.32 | -1.24 | 0.27 | -0.33 | -1.30 | -3.92 |
| **Ger1(cw)** | 0.89 | -1.77 | -0.61 | -1.11 | -0.64 | -1.50 | 0.91 | -2.37 |
| **Ger2(dh)** | -0.30 | -2.76 | -0.08 | -1.37 | -0.92 | -1.50 | -0.38 | -1.09 |
| **Ger3(js)** | -0.53 | -3.10 | -0.08 | -1.45 | -0.20 | -1.73 | 1.82 | -0.14 |
| **Hin1(k2)** | -0.84 | -2.96 | -0.34 | -1.86 | -0.95 | -1.66 | 0.25 | -2.64 |
| **Tam1(s4)** | -1.33 | -2.86 | -0.16 | -2.18 | -1.46 | -2.53 | -0.60 | -2.85 |

TABLE I: *Mean z-scored vocal tract area descriptors (VTADs) A1, A2, A3+A4 (denoted as A3) and Jaw angle (JA) for postures assumed on average during (a) inter-speech pauses (I) and (b) absolute rest positions (R). See Figure 1 for details on VTADs.*

| Speaker | JA | TL | LA | TTCD | TDCD | TRCD | VEL |
|---|---|---|---|---|---|---|---|
| **Eng1(m1)** | 0.0137 | 0.0289 | 0.0814 | 0.1780 | 0.0547 | 0.0985 | 0.4844 |
| **Eng2(s2)** | 0.0118 | 0.0227 | 0.0596 | 0.3243 | 0.1436 | 0.1036 | 0.2393 |
| **Eng3(l4)** | 0.0345 | 0.0431 | 0.2460 | 0.2706 | 0.5503 | 0.1608 | 0.6358 |
| **Eng4(h5)** | 0.0608 | 0.0187 | 0.0705 | 0.2042 | 1.2054 | 0.2733 | 0.2934 |
| **Eng5(d6)** | 0.0454 | 0.1466 | 0.5547 | 0.6853 | 0.3274 | 0.1259 | 0.5664 |
| **Ger1(cw)** | 0.0428 | 0.0359 | 0.2190 | 0.2197 | 0.2218 | 0.1628 | 0.6525 |
| **Ger2(dh)** | 0.0364 | 0.0398 | 0.5528 | 0.3247 | 0.6177 | 0.1731 | 0.5738 |
| **Ger3(js)** | 0.1161 | 0.1360 | 0.9401 | 0.8734 | 0.4986 | 0.2227 | 0.7480 |
| **Hin1(k2)** | 0.0936 | 0.1346 | 0.6122 | 1.4058 | 0.5039 | 0.1376 | 0.9620 |
| **Tam1(s4)** | 0.0765 | 0.1064 | 0.4428 | 0.4805 | 0.4972 | 0.4006 | 1.1376 |

TABLE II: *Coefficients of variance of (1) task variables - jaw angle (JA) and tongue length (TL), and (2) articulator variables - lip aperture (LA), tongue tip, dorsum & root constriction degrees (TTCD, TDCD & TRCD) and velic aperture (VEL) for all 10 speakers.*

sum and velum). Calculated on the MR image of the midsagittal plane, vocal tract area descriptors (VTADs) are computed as the areas of these regions (for example, A1 roughly approximates area of the region enclosed from the lips to the tongue tip constricting device, A2 – from the tongue tip to dorsum, and A3 – tongue dorsum to pharynx). See Figure 1. For more details, please see [2]. To obtain articulator-based measures, we measure the jaw angle as the obtuse angle between regression lines fitted to the pharyngeal wall and the contour of the jaw, and length of the contour of the tongue in the midsagittal image.

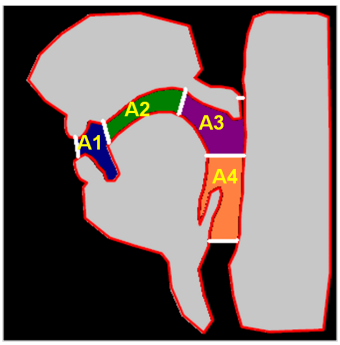


Fig. 1: *A schematic of vocal tract area descriptors (VTADs). VTAD A1 roughly approximates area enclosed from the lips to the tongue tip, A2 – from the tongue tip to dorsum, and A3 – tongue dorsum to pharynx. JawAngle is the obtuse angle between the pharyngeal wall and a regression line fitted to the jaw). For more details, see [2].*

Our initial results are as follows: for speakers of *all* language backgrounds, we see a significant difference between postures assumed during inter-speech postures and absolute rest positions, i.e., the vocal tract is more constricted, with a lower jaw height for absolute rest positions (see Table I). This is consistent with previously obtained results [2].

We also attempt to understand the nature and representations of control of these articulatory settings in this paper. Working within the framework of Task Dynamics [1], we want to find which control specification is more likely for these ASs – a articulator-based control or a tract(or task)-variable-based control. In order to answer this question, we compute the coefficients of variability (COV) of both articulatory variables (such as the total length of the tongue and jaw angle) and task variables (such as lip aperture, tongue-tip constriction degree, tongue-dorsum constriction degree, tongue-root constriction degree and velum aperture; see [1] for an in-depth description of these task variables), and compare these COV values for each of the 10 speakers using pair-wise statistical tests. We observe a significantly *higher* COV for task variables (see Table II) over these intervals than for articulatory variables (significant for all cases tested) especially in native English, which lends support to the hypothesis that articulator kinematics (like the position and velocity of the jaw) are the variables controlled in order to specify the target postures observed in different settings, rather than tract variable kinematics (like magnitude and speed of lip aperture, etc). [Supported by NIH]

## REFERENCES

[1] E. Saltzman and K. Munhall, "A dynamical approach to gestural patterning in speech production," *Ecological Psychology*, vol. 1, no. 4, pp. 333–382, 1989.

[2] V. Ramanarayanan, D. Byrd, L. Goldstein, and S. Narayanan, "Investigating articulatory setting – pauses, ready position, and rest – using real-time MRI," *Interspeech 2010, Makuhari, Japan*, 2010.

[3] S. Narayanan, K. Nayak, S. Lee, A. Sethy, and D. Byrd, "An approach to real-time magnetic resonance imaging for speech production," *The Journal of the Acoustical Society of America*, vol. 115, p. 1771, 2004.

[4] B. Gick, I. Wilson, K. Koch, and C. Cook, "Language-specific articulatory settings: Evidence from inter-utterance rest position," *Phonetica*, vol. 61, pp. 220–233, 2004.