Functional segments in tongue movement

Maureen Stone*
Melissa A. Epstein*
Khalil Iskarous[+]

*Vocal Tract Visualization Lab, University of Maryland Dental School, Baltimore, Maryland, USA
[+]Haskins Lab, Yale University, New Haven, Connecticut, USA

Address correspondence to:
Dr. Maureen Stone
Department of Biomedical Sciences, Rm. 5A12
University of Maryland Dental School
666 W. Baltimore St.
Baltimore, MD 21201
USA

Tel: ++410-706-1269
Fax: ++410-706-0193

mstone@umaryland.edu

Key words: function segments, MRI, tongue, ultrasound

Abstract

The tongue is a deformable object, and moves by compressing or expanding local functional segments. For any single phoneme, these functional tongue segments may move in similar or opposite directions, and may reach target maximum synchronously or not. This paper will discuss the independence of five proposed segments in the production of speech. Three studies used ultrasound and tagged Cine-MRI to explore the independence of the tongue segments. High correlations between tongue segments would suggest passive biomechanical constraints and low correlations would suggest active independent control. Both physiological and higher level linguistic constraints were seen in the correlation patterns. Physiological constraints were supported by high correlations between adjacent segments (positive) and distant segments (negative). Linguistic constraints were supported by segmental correlations that changed with the phonemic content of the task.

Introduction

In the speech production process, the tongue deforms in a complex fashion, often executing the gestures for two or more phonemes simultaneously. A basic question in speech production research is how the highly deformable tissue of the tongue is controlled for the accomplishment of speech. A wide range of hypotheses about the basic control units of tongue movement has been entertained in the speech production literature. Early work on lingual coarticulation (Hardcastle, 1976, 1985, Ohman, 1967, Mermelstein, 1973) postulated the tongue to be divided into tip and body executing quasi-independent motions. Geometric models of the tongue also assumed this division (cf. Coker and Fujimura, 1965; Mermelstein, 1973), as did articulatory synthesis (Rubin, 1981) which produced spectra closely matching human speech. Factor analytic models found two or three basic factors that captured tongue contour shape, especially for vowels. The full range of tongue shapes were then modeled as linear combinations of these basic factors (cf. Harshman, Ladefoged and Goldstein, 1977; Maeda, 1989; Stone, Goldstein and Zhang, 1997; Slud, Stone, Smith and Goldstein, 2002). More recently, however, there has been mounting evidence that higher dimensional representation for the tongue is warranted. For example, imaging techniques have allowed multiple tongue planes and clearer tongue contours to be visualized. These data are then used in three-dimensional surface reconstructions, which show that tongue surface deformations are far more complex than the motion of a tongue body and a tongue tip (cf. Wilhelms-Tricarico, 1995; Stone and Lundberg, 1996; Yang and Stone, 2002). Here we pursue a different hypothesis. We view the tongue as composed of a number of coronal segments that are quasi-independently controlled.

Evidence for this hypothesis comes from a variety of sources. Neuroanatomical studies have indicated potentially very complex innervation of the tongue. A number of studies indicate that the human hypoglossal nucleus is composed of at least 6500 motoneurons per side (Wozniak and Young, 1969; Atsumi and Miyatake, 1987; O'Kusky and Norman, 1995); thus at least 13,000 hypoglossal motoneurons innervate the human tongue. For example, the tongue contains repeating, alternating, laminae of vertical (genioglossus(GG), verticalis (V)) and horizontal (transverses(T)) fibers, each of which contains its own motoneuron (Mu and Sanders, 1999; Takemoto, 2001). Since every motoneuron is potentially an independent unit of control, then T, V and GG could have multiple segments of control (cf. Miyawaki, 1975). Tongue neuronantomy therefore provides evidence for movement by local control along the tongue, as opposed to control of the tongue body as a unit.

It is unlikely however that each motor unit, or small number of units, is separately controlled in the speech production task. Several theories have proposed mechanisms whereby the control system reduces the number of degrees freedom at its disposal. Action theory approaches to peripheral control have modeled rapid complex shape changes using synergistic muscle actions or coordinative structures (Turvey, 1977; Fowler, 1980). In that model, muscles behave as synergistic groups to create specific shapes. The synergies or structures act as a single unit with trade-offs in the amount of contraction for any single muscle in the group. This allows a simpler control mechanism than one in which each muscle independently receives a command for each gesture.

We aim for an intermediate model with a larger number of degrees of freedom than a two-part tongue (tip and body), but far fewer than 13,000. Our hypothesis is that tongue deformation is controlled by the synergistic coordination of 'functional segments' of the tongue. These segments are laid out orthogonally to the longitudinal axis of the vocal tract as portrayed in figure 1 and composed of multiple muscle systems. This paper is a first step in determining how many independent segments comprise the tongue and how they relate to its muscular architecture.

Figure 1 about here

Tongue muscle architecture is complex and contains three interesting features that allow us to consider functional segments as the units of control. First, the tongue is anisotropic in the anterior-posterior direction. That is, the muscles of the tongue differ from front-to-back in fiber size, and more importantly, direction. This is especially notable for styloglossus (SG) and hyoglossus (HG). Second, the tongue is anisotropic from medial-to-lateral. The extrinsic tongue muscles are either very medial (GG) or very lateral (SG, HG). Third, the tongue's two sides are mirror images of each other, and can move symmetrically or in rotation (Stone, 1990). Considering the complex neuroanatomy and motion of the tongue, along with the possibility of controlling that complexity with coordinated muscle aggregates, we believe that the tongue can be divided into quasi-independently controlled functional segments based on regions of the tongue and vocal tract, rather than gross muscle architecture. Instead of entire muscles aligning to execute a gesture, segments would be controlled independently or aggregated into larger units to form coordinative structures determined by language dependent phonetic considerations.

An example of segmental control is found in retroflexion, which is better explained using functional segments than whole muscle activity. Tongue-tip retroflexion, as seen on MRI (Narayanan, Alwan and Haker, 1997; Ong and Stone, 1998), would be hard to produce using the most likely whole muscle strategy, i.e. contraction of superior longitudinalis (SL) to bend the tongue and of genioglossus anterior (GGA) to create an anterior depression. However, the anatomy of SL may be better suited to retroflexion than previously thought. In humans, SL is not composed of parallel fibers. Rather it is a muscle of in-series design, in which short muscle fibers overlap to extend from muscle origin to muscle insertion (Sokoloff, 2003). From studies of the SL in the rat (Sokoloff, 2000) it is likely that SL motoneurons in the human project to limited antero-posterior regions of the SL system, thus providing a mechanism for a localized innervation of SL fibers. If so, in-series fibers of SL could be activated exclusively at the tip to facilitate retroflexion. Figure 2 depicts two deformations of the upper surface of the tongue. The images are from a finite element model (FEM) of the outer muscles of the tongue in an unbent state, i.e. shaped like an uncurved rectangular solid. The FEM has 5 segments that can be controlled independently (Essex-Torcaso and Levine, 2003). Figure 2 shows the output of this model. Both figures contract GG and V in segment 1 (tongue tip). SL is contracting in all five segments (left) or only in segment 1 (right). Clearly, contraction of SL in the tip provides more realistic retroflexion.

Figure 2 about here

The large number of motoneurons, the complex fiber orientation and fiber type, coupled with the idea of coordinative structures, have encouraged us to consider tongue segments that transcend specific muscles. Two studies were designed to explore the number and composition of tongue segments that might be independently controlled. The first one used ultrasound to correlate midline tongue height at different tongue segments. The second used tagged Cine-MRI (tMRI) to compare compression and expansion patterns at different tongue segments consistent with the lines of action of several muscles.

The ultrasound study consisted of two experiments that examined midsagittal and coronal data to determine the extent to which different parts of the tongue are independently controlled by looking at the correlation in the movement at different sections. High correlation between two segments indicates the two segments act as a unit, whereas low correlation is indicative of independence of control. Previous work showed that five to six optimized midsagittal points could reconstruct the midline tongue contours of the US English vowels /i, ɑ, æ, u/ with very small error (Lundberg and Stone, 1999). Therefore, the first experiment looked at correlations between five midsagittal points, optimized in location from front to back for phonetically balanced sentences. The second experiment looked at the issue in more depth by examining two sentences that varied in a single phoneme.

The tMRI study explored how functional segments might be organized in terms of muscle activity. Since muscles can only contract, local contraction of a muscle would cause local compression, and both local and distal expansion (orthogonal to compression). To determine the local and distal relationships between tongue segments for different speech materials we examined data from tagged Cine-MRI (tMRI). tMRI allows tracking of tissue points on and within the tongue from which compression and expansion can be calculated. In the present study we considered the agonist/antagonist relationship between GG, V and T and how that relationship differed at different regions of the tongue and for different speech tasks. In other words, T and V are usually considered to be single muscles. The goal was to see whether unitary compression would be visible, or whether local tradeoffs between GG, V and T would be consistent with functional segments.

It should be noted that the segments in the four datasets are not identical due to different methods of data analysis and different subjects. However, they are fairly close to each other (within a few mm) and the results suggest they represent comparable segments.

Ultrasound Study,

Experiment 1. Correlation of midline height at 5 surface locations, for a phonetically balanced speech sample.

Study 1 examined the independence of local tongue segments in the production of speech using ultrasound imaging. The goal was to explore the movement correlations in five tongue segments (see figures 1 and 3). High correlations among the segments would indicate dependence, i.e. lack of functional segments. Low correlations would support local independence. A large number of independent segments would imply a more complex motor control strategy, i.e. independent control of tip, blade, etc.

Figure 3 about here.

Experiment 1 tested the independence of tongue segments using five of the fourteen sentences in The Fisher Logemann Test of Articulatory Competence (1971). The test is designed to measure articulation clarity, and each sentence is loaded with a specific type of sound, e.g. sentence 2 contains many /dʒ/ and /tʃ/ productions. These particular sentences were chosen because in combination they are phonetically balanced. We investigated the correlations in the data as a whole, and in individual sentences. The use of phonetically balanced data is unusual in speech production research, but it is essential in this study for interpreting the correlational data. We hypothesize that the correlations in the data have two sources: physiological and phonetic. Two segments could be highly correlated (positively or negatively) due to the physical link between proximal and distal tongue sections. Alternatively, they could be highly correlated due to language specific constraints. By comparing the correlations in the phonetically balanced data and the individual sentences, we should be able to separate the physical and phonetic contributions to the correlations.

Speaker 1 repeated the sentences once each, while seated in the HATS recording system, and midsagittal ultrasound scans were collected (Stone and Davis, 1995). Midsagittal tongue shape was determined by calculating midline displacement at five points along the midsagittal slice (figure 3-left side). These five points were determined by an optimization programme that chooses the five points that best reconstruct the midsagittal curve (Lundberg and Stone, 1999). Pearson product moment correlation coefficients were calculated for all combinations of the five points. Correlations of $r = 0.70$ or greater (accounting for 49% or more of the variance) were considered ‘strong’.

Results and Discussion

Several interesting results emerged from this data set (see table 1). Strong positive correlations occurred between adjacent segments, consistent with close physical coupling. All strong negative correlations occurred between the anterior and posterior regions, with segment 3 being the dividing point. Segment 3 had only one strong correlation (segment 4, sentence 1). This behavior is consistent with a tongue that has two major regions of activity, front and back, separated at segment 3, and hydrostatic structure that links movement of the front to the back: when the oral portion of the tongue is forced down, the pharyngeal portion moves towards the back, and vice versa. There also were no strong correlations between segments 2 and 4.

Table 1 about here

Close examination of the correlations in the individual sentences (table 1), revealed linguistic effects in the segmental behavior. First, tongue segment size differed from sentence to sentence. In sentence two, segments 4 and 5 were both correlated with segment 1; i.e. they acted as one segment. In sentence three segments 1 and 2 acted as one segment in that both correlated highly with segment 5. This difference indicates that local segments can group into larger regions of motion, i.e. two segments can act as one, but do not have to. Second, some segmentation appears to be phoneme dependent. The most apparent of these is in sentence two. Segments 2 and 3 are strongly correlated in every sentence but this one. This lack of correlation where there was typically a strong one, suggests that the predominance of palatal sounds in sentence two decoupled segments 2 and 3. Similarly, sentence two has coupled segments 1 and 4, as it is the only sentence with a strong negative correlation between them. Thus, although the most salient results are that of an anterior/posterior tradeoff, the more interesting result, that of variable segmental correlations, advocates for local planning. To test this further, experiment 2 considered sentences that varied only in one phoneme.

Experiment 2. Correlation of midline heights for two sentences.

The second ultrasound experiment considered tongue height in a controlled phonetic dataset. The dataset, described in full elsewhere (Yang and Stone, 2002), contains coronal ultrasound images of the sentences ‘It rang a lot’, and ‘It ran a lot’. Speaker 2 repeated the sentences once each for data collection at five coronal slices. The five image-sequences were temporally aligned using the acoustic wave so that comparable ultrasound frames could be compared. The data sets were collected in the coronal plane at the radial angles depicted on the right side of figure 3. The midline x-value was determined by overlaying all 90 coronal contours; midline y-displacement (hereafter: height) was then recorded at each midline x-value. Correlations were done comparing all five segments.

Results and Discussion

In general, this data set showed the same effects as experiment 1. First, there was evidence of tongue segment coupling. Table 2 indicated ten strong height correlations of 30 (ten each for ran, rang, and combined data). For ran, segments 1 and 2 correlated strongly with 3. For rang, 1 and 2 correlated with 3 and 4. We interpret this to mean that the segments combined to form larger segments 1+2 and 3+4, reducing the number of degrees of freedom of the tongue by making these adjacent segments functionally dependent. Second, segmental correlations were phoneme dependent. There were more correlations for rang than ran. The velar gesture appeared to dominate a large part of the tongue motion sequence. Finally, both sentences had a division between slices 2 and 3. Slices on either side correlated positively with each other. Slices across the divide correlated negatively. In other words, slices 1,2 opposed slices 3,4,5. In particular, the strong correlations for rang are consistent with the tongue having two major segments, a tip/blade and body. The relative lack of correlation in ran suggests a difference in segmental organization between rang and ran, in which ran does not use segments 4,5 in correlation with 3.

In the ultrasound study we have investigated the correlations in the movements at different sections of the vocal tract. This is also the first step in factor-analytic studies of tongue movement (Ladefoged, et al., 1977; Sanguineti, Laboissiere and Ostry, 1998; Maeda, 1989). The data for factor analysis studies consist of either static tongue surface shapes (Ladefoged, et al., 1977; Maeda, 1989) or dynamic tongue surface motion (Sanguineti, et al., 1998). Covariance matrices are constructed for the data, which is basically the same as the correlational analysis in the current study. The next step in the factor-analytic technique is to extract the principal components (eigenvectors) of the covariance matrix. The first two or three principal components (if their eigenvalues are high) indicate the vocal tract segments where most of the variation occurs. As we saw earlier, however, the correlation between segments can be due to either a physical link or a phonetic link. Factor-analytic techniques extract the segments showing the greatest correlation, but do so only from the pooled data set, so are unable to compare the physical and phonetic aspects of the correlation. Examination of the overall shape of the first two factors in most of the factor-analytic studies shows a general front-back factor, which is consistent with the hydrostatic nature of the tongue. Correlational studies for the investigation of the motor control of the tongue require however that the phonetic component of the correlation be separated out and investigated.

Table 2 about here

In sum, the ultrasound study supports an overall division of the tongue into two large regions in which adjacent segments correlate positively, and distant ones negatively, with a division around segment 3 (see figure 3). However, individual correlations indicate that segments can be coupled for certain motions and that the correlations and couplings are phoneme dependent. Similar results have been found for the temporal aspects of coarticulation (cf. Gibbon et al., 1993). Additional ultrasound data are needed to strengthen this finding. To explore this notion further, it is necessary to consider possible muscle activity. One way to do this is through tMRI.

Tagged Cine-MRI (tMRI) Study. Examination of muscle trade-offs in functional segments.

Whereas ultrasound imaging detects motion of the tongue surface at fixed sections of the vocal tract (figure 3), tMRI detects motion of tissue points on the surface and internal to the tongue (figure 4). Therefore, it can be used to infer muscle contraction. In addition, ultrasound and tMRI data can be coregistered to the same coordinate space (Stone, Parthasarathy, Iskarous, NessAiver and Prince, 2003). Ultrasound shows that the tongue behaves as if it has segments. tMRI addresses how these segments are structured. A tMRI study was done to explore the functional linkages between the muscles underlying the five segments. Specifically we studied the muscles genioglossus (GG), transversus (T) and verticalis (V). As mentioned earlier, these muscles are organized into hundreds of alternating lamina, composed of vertical and horizontal fibers, with separate motoneurons. It is worthwhile, therefore, to examine differences in the local behavior of these muscles.

It is known that GG has an architecture that could allow local activation at different segments (from genioglossus anterior (GGA) to genioglossus posterior (GGP)) (Miyawaki, 1975). The question is whether T and V are comparable. There are three possible patterns for both T and V. (1) They do not show GG-like segmental behaviours. (2) They show segmental behaviors that are highly correlated with GG. (3) They show segmental behaviors that are not correlated with GG. To test this we examined muscle compression from tMRI data for /ʃɑ/, /kɑ/ and /ku/ to see how different tongue segments behave in different consonant-vowel sequences.

MRI measures the hydrogen content of biological tissue. Since both water and fat have a high hydrogen content, tissue high in water or fat appears more white on an MR image. Tissue without water, such as bone or teeth, appear black. While standard MRI visualizes tissue types well, neither it, nor ultrasound, can identify specific tissue points. tMRI overcomes this problem. Tags are created in tissue by demodulating the spin phase of some of the hydrogen protons. The demodulation occurs in alternating planes, and the demodulated protons are invisible in the scanned image. Thus tMR images have black stripes on them which are slices of tissue that have not been imaged. When a structure (such as the tongue) moves, the tagged portions move too and tissue deformation can be seen (see figure 4). From motion of the tissue, compression can be calculated anywhere in the tongue, including along the line of action of a specific muscle. However, since some compression is passive, muscle contraction can only be inferred, not proven, from such compression.

Figure 4 about here.

Methods:

Two new subjects were used. Subject 3, a 23 year old male, native speaker of US English (USE) repeated the syllable /ʃɑ/ at a rate of once per second. Twenty four time-phases (frames), at 41ms each, were measured for each repetition. Ten sagittal and ten coronal tongue slices were measured, with 13 repetitions of the utterance required per slice. Multiple repetitions are required because hydrogen protons give off weak signals and multiple scans are summed into one composite image. Slice thickness was 7 mm and tag decay time 600 ms. The protocol used a sound gated standard SPGR cine on a 1.5T GE Signa Scanner (TR = 6.2, TE = 1.65, FA = 10, 256 x 128 matrix, FOV = 30cm). This data set is discussed in Dick, Ozturk, Douglas, McVeigh and Stone (2000). Subject 4, a 19 year old male native speaker of USE repeated the syllables /kɑ/ and /ku/ once per second, while tMRI images were collected in three sagittal planes. This data set is discussed in detail in Stone, Davis, Douglas, NessAiver, Gullapalli, Levine and Lundberg (2001, a; b). Although subjects 3 and 4 had slightly different methods, their results overlap and are discussed together. For both subjects, stretch (compression/expansion) was measured at five front-to-back locations for GG, V-left and –right, (and T for subject 3) (see figure 5). For subject 3 the data were combined into a 3D volume and the 3 muscles were measured according to their fiber directions, origins and insertions at each segment. Because subject 4's data contained only sagittal slices, T could not be measured; V was measured from the left and right sagittal slices, and GG from the midsagittal slice. Statistical analyses were not conducted due to the small size of the data set.

Figure 5 about here

Results and Discussion.

Stretch for the three muscles at the five tongue locations, is shown over time for /ʃɑ/ (see figure 6). A comparison of GG and T shows several relationships. First, at each segment GG and T mirrored each other in magnitude. There was graded compression from anterior-to-posterior for GG (left) and graded expansion from anterior-to-posterior for T (middle) culminating in the maximum /ɑ/ position (arrows). The GG compression pattern was consistent with a single contraction occurring at the anteriormost segment (1) and propagating to neighboring segments (2, 3), which showed less compression. Second, the compression of GG1 caused a great expansion in T. Since the tongue motion from /ʃ/ to /ɑ/ is considered to be downward and backward, it was expected that the major directions of stretch would be compression in GG1 and expansion in GG5. GG1 was indeed the most compressed (black diamond), but, GG5 (white squares) was not very expanded, T1 expansion (black diamonds) indicated a local lateral expansion (see also figure 4).

Figure 6 about here

Although tMRI cannot definitively distinguish active from passive compression, it is interesting to note several differences between the passive expansion of T1 and the possibly active compression of GG1. The slope of GG1 compression was steep, then flat; the expansion of T1 was more linear. Perhaps active compression is rapid, whereas passive compression would be slower, more similar to expansion (which is always passive). Also, T1 started to compress before GG1 started to expand (arrows). The time lag may indicate active contraction of T1 (or another muscle), which initiates the movement back to /ʃ/ position.

These observations lead to some ideas about tongue control. First, these two antagonists (GG, T) appear to be a functional as well as a structural unit for the /ʃɑ/ gesture. Second, the three anterior segments act in opposition to the two posterior ones for both muscles. The notion of multiple activations is well-accepted for GG, but unexplored for T. These data suggest that T has at least two functional segments (anterior, posterior). Interestingly, V did not act in concert with GG and T. All segments of the muscle compressed in unison (segment 5 oscillations are noise). This is consistent with the traditional representation of V as a single unit. Since these results were based on a single subject and syllable, corroboration was sought in the data from subject 4.

Subject 4 spoke the syllables /kɑ/ and /ku/ while left-, mid-, and right sagittal slices were collected on tMRI. Figure 7 shows percent stretch for (a) V-left, (b) GG, (c) V-right at the five tongue segments shown in figure 5. Similarities can be seen between /kɑ/, /ku/ and /ʃɑ/ (/ʃɑ/ is spoken by subject 3). For each syllable, compression occurred in tongue segments 1 and/or 2, although the compressed segment(s) differ for each syllable. Considering the midline motion first, GG showed three, phonemically related, compression patterns at the vowel maximum (arrows). For /ku/, segment 2 was more compressed than 1, creating a high-back to less high-back motion. For /kɑ/, 1 and 2 were equally compressed, that is, coupled to create a high-back to low-back motion. For /ʃɑ/ 1 was more compressed than 2, creating a high-front to low-back motion. This also may reflect a subject effect since /ʃɑ/ was spoken by subject 3, and /kɑ/ and /ku/ by subject 4.

Figure 7 about here

Recall that for subject 3, V showed unitary contraction from /ʃ/ to /ɑ/. For subject 4, V differed left-to-right and front-to-back. Left-to-right asymmetry reflected tongue rotation. In both /kɑ/ and /ku/, one side had large compression for segment 2, but not 1, indicative of a stable pivot point at segment 1. For /ku/ the pivot was on the left, and for /kɑ/, on the right. Previous observation indicates that a single subject is often asymmetrical on different sides for different syllables (cf. Stone, 1990). On the opposing side, V compressed at both segments 1 and 2, consistent with moving more mass on that side. Whether the differences in V between /k/ and /ʃ/ are due to subject or task need to be studied further.

The tMRI data, alternating trading-offs in compression and expansion between the anterior and posterior tongue, supported the ultrasound correlations dividing the tongue into two large regions at about segment 3. The tMRI data also showed segment like behavior for V and T, similar to GG (subject 4) in that their anterior and posterior segments moved in opposition, and local segmental behavior was evident across syllables as well. T and GG were more tightly coupled than V (subject 3). Thus these three muscles appear to have functional segments strongly affected by phonemic constraints. It was also seen that V can use one-sided or weighted activation patterns to create left-to-right asymmetries; such rotation is an economical method for producing the rapid, large motions required for these syllables. Finally, the data revealed that volume shifting can be greater locally than distally. This was seen in the large trade-off between GG1 and T1 (local), compared to a smaller trade-off between GG1 and GG5 (distal).

Conclusions

This study asked whether the tongue is composed of two parts, a body and tip, or multiple functional segments. Evidence for functional segments was found in both datasets and for all four subjects. The ultrasound correlations and the tMRI compression patterns showed local phonemic effects. The data also showed that the segments could act in unison, simulating a 'tongue body'. This is not evidence for the earlier notion that the tongue is simply a body and a tip, since the functional segments may or may not coordinate their action to result in an apparently unitary tongue body. Further data are needed to strengthen these results.

Acknowledgements

The authors would like to thank Alan Sokoloff for his very helpful comments and discussion. The project was supported in part by grant no. R01-DC01578 from the National Institute of Deafness and other Communication Disorders, NIH.

References

Atsumi T., and Miyatake T., 1987, Morphometry of the degenerative process in the hypoglossal nerves in amyotrophic lateral sclerosis. *Acta Neuropathologica*, **73**, 25-31.

Coker, C.H., and Fujimura, O., 1966, Model for specification of vocal-tract area function. *Journal of the Acoustical Society of America*, **40**, 1271.

Dick, D., Ozturk, C., Douglas, A., McVeigh, E., and Stone, M., 2000, Three-dimensional tracking of tongue motion using tagged MRI. *Proceedings of the International Society for Magnetic Resonance in Medicine, 8th Scientific Meeting and Exhibition*, 1-7 April, 2000, Denver, Colorado.

Essex-Torcaso, C., and Levine, W.S., 2003, A simplified finite element model of the human tongue. Technical Report TR2003-35, ISR, University of Maryland, College Park.

Fisher, H. B and Logemann, J. A., 1971, The Fisher-Logemann Test of Articulation Competence, (Boston: Houghton Mifflin).

Fowler, C., 1980, Coarticulation and theories of extrinsic timing control. *Journal of Phonetics*, **8**, 113-133.

Gibbon, F., Hardcastle, W, and Nicolaides, K, 1993, Temporal and spatial aspects of lingual coarticulation in /kl/ sequences: A cross-linguistic investigation. *Language and Speech*. **36**, 261-277.

Hardcastle, W.J., (1976) *Physiology of Speech Production*. London: Academic Press.

Hardcastle, W.J., (1985) Some phonetic and syntactic constraints on lingual coarticulation during /kl/ sequences. *Speech Communication*, **4**, 247-263.

Harshman, R. A., Ladefoged, P., and Goldstein, L., 1977, Factor analysis of tongue shapes. *Journal of the Acoustical Society of America*, **62**, 693-707.

Lundberg, A., and Stone, M., 1999, Three-dimensional tongue surface reconstruction: Practical considerations for ultrasound data. *Journal of the Acoustical Society of America*, **106**, 2858- 2867.

Maeda, S., 1991, On articulatory and acoustic variabilities. *Journal of Phonetics*, **19**, 321-331.

Mermelstein, P., 1973, Articulatory model for the study of speech production. *Journal of the Acoustical Society of America*, **53**, 1070-1080.

Miyawaki, K., 1975, A preliminary report on the electromyographic study of the activity of lingual muscles. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics, University of Tokyo*, **9**, 91–106.

Mu, L., and Sanders, I., 1999, Neuromuscular organization of the canine tongue. *The Anatomical Record*, **256**, 412-424

Narayanan, S., Alwan, A., and Haker, K., 1997, Towards articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part 1: The laterals. *Journal of the Acoustical Society of America*, **101**, 1064-1077.

Ohman, S.E.G., 1967, Numerical model of coarticulation. *Journal of the Acoustical Society of America*, **41**, 310-320.

O'Kusky, J.R., and Norman, M.G., 1995, Sudden infant death syndrome: increased number of synapses in the hypoglossal nucleus. *Journal of Neuropathology and Experimental Neurology*, **54,** 627-34.

Ong, D., and Stone, M., 1998, Three-dimensional vocal tract shapes in [r] and [l]: A study of MRI, ultrasound, electropalatopgraphy, and acoustics. *Phonoscope*, **1**, 1-14.

Rubin, P., Baer, T. and Mermelstein, P., 1981, An articulatory synthesizer for perceptual research. *Journal of the Acoustical Society of America*, **70**, 321-328.

Sanguineti, V., Laboissiere, R., and Ostry, D.J., 1998, A dynamic biomechanical model for neural control of speech production. *Journal of the Acoustical Society of America*, **103**, 1615-1627.

Slud, E., Stone, M., Smith, P.J., and Goldstein, M., 2002, Principal components representation of the two-dimensional coronal tongue surface. *Phonetica*, **59**, 108-133.

Sokoloff, A.J., 2000, Localization and contractile properties of intrinsic longitudinal motor units of the rat tongue. *Journal of Neurophysiology*, **84**, 827-835.

Sokoloff, A.M., 2003, Neuromuscular strategies for human tongue movement. Proceedings of the Fifteenth International Congress of Phonetic Sciences. 3-9 August, Barcelona, Spain, pp. 3181-3183.

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. and Davis, E.P., 1995, A head and transducer support system for making ultrasound images of tongue/jaw movement. *Journal of the Acoustical Society of America*, **98,** 3107-3112.

Stone, M., Davis, E., Douglas, A., NessAiver, M., Gullapalli, R., Levine, W., and Lundberg, A., 2001a, Modeling motion of the internal tongue from tagged cine-MRI images. *Journal of the Acoustical Society of America*, **109**, 2974-2982.

Stone, M., Davis, E., Douglas, A., NessAiver, M., Gullapalli, R., Levine, W., and Lundberg, A., 2001b, Modeling tongue surface contours from cine-MRI images. *Journal of Speech, Language, and Hearing Research*, **44**, 1026-1040.

Stone, M., Goldstein, M., and Zhang, Y., 1997, Principal component analysis of cross-sectional tongue shapes in vowels. *Speech Communication*, **22**, 173-184.

Stone, M., and Lundberg, A., 1996, Three-dimensional tongue surface shapes of English consonants and vowels. *Journal of the Acoustical Society of America*, **99**, 3728–3737.

Stone, M., Parthasarathy, V., Iskarous, K., NessAiver, M., and Prince, J., 2003, Tissue strains and tongue shapes: combining tMRI and ultrasound. Proceedings of the Fifteenth International Congress of Phonetic Sciences. 3-9 August, Barcelona, Spain, pp. 273-276.

Wilhelms-Tricarico, R., 1995, Physiological modeling of speech production: Methods for modeling soft-tissue articulators. *Journal of the Acoustical Society of America*, **97**, 3085–3098.

Takemoto, H., 2001, Morphological analyses of the human tongue musculature for three-dimensional modeling. *Journal of Speech and Hearing Research*, **44**, 95-107.

Turvey, M.T., 1977, Preliminaries to a theory of action with reference to vision. In: R. Shaw & J. Bransford (eds) *Perceiving, Acting and Knowing: Toward an Ecological Psychology*, (Hillsdale, NJ: Lawrence Erlbaum Associates), pp. 211-265.

Wozniak W. and Young P.A., 1969, Further observations on human hypoglossal nerve. *Anatomischer Anzeiger*, **125**, 203-5.

Yang, C.S., and Stone, M., 2002, Dynamic programming method for temporal registration of three-dimensional tongue surface motion from multiple utterances. *Speech Communication*, **38**, 199-207.

Table 1. Correlations between segments for the Fisher-Logemann sentences, combined and separately. Bolded values are strong at r≥0.70 and significant at p≤0.001.

All Sentences (n=466 contours)

	Segment 2	Segment 3	Segment 4	Segment 5
Segment 1	0.65	0.16	-0.47	-0.59
Segment 2		**0.73**	-0.20	-0.69
Segment 3			0.37	-0.24
Segment 4				**0.70**

1. The girls were baking the biggest cake for Mr. Tag. (n=86 contours)

	Segment 2	Segment 3	Segment 4	Segment 5
Segment 1	0.35	-0.06	-0.21	-0.18
Segment 2		**0.76**	0.37	-0.41
Segment 3			**0.83**	0.06
Segment 4				0.49

2. George is at the church watching a magic show. (n=79 contours)

	Segment 2	Segment 3	Segment 4	Segment 5
Segment 1	**0.71**	-0.19	**-0.71**	**-0.70**
Segment 2		0.47	-0.34	-0.55
Segment 3			0.58	0.22
Segment 4				**0.85**

3. Nancy found some fine hangers among the many things at the sale. (n=106 contours)

	Segment 2	Segment 3	Segment 4	Segment 5
Segment 1	**0.79**	0.56	-0.40	**-0.72**
Segment 2		**0.89**	-0.20	**-0.81**
Segment 3			0.20	-0.59
Segment 4				0.59

4. Let me keep a little of this wedding cake to eat later. (n=78 contours)

	Segment 2	Segment 3	Segment 4	Segment 5
Segment 1	0.68	0.47	-0.44	-0.64
Segment 2		**0.89**	-0.28	**-0.88**
Segment 3			0.11	-0.65
Segment 4				0.55

5. Ruth caught a cold because she wouldn't wear her new warm wool coat. (n=117 contours)

	Segment 2	Segment 3	Segment 4	Segment 5
Segment 1	**0.71**	0.22	-0.59	-0.61
Segment 2		**0.73**	-0.60	**-0.84**
Segment 3			-0.26	-0.58
Segment 4				**0.88**

Table 2. Correlations between segments for the ran/rang sentences. Bolded values are strong at r≥0.70 and significant at p≤0.001.

It ran a lot. (n=45 contours).

	Segment 3	Segment 4	Segment 5	Segment 6
Segment 2	0.66	**-0.83**	-0.36	0.06
Segment 3		**-0.76**	-0.24	0.27
Segment 4			0.30	-0.07
Segment 5				0.66

It rang a lot. (n=45 contours).

	Segment 3	Segment 4	Segment 5	Segment 6
Segment 2	0.59	**-0.90**	**-0.70**	-0.43
Segment 3		**-0.70**	**-0.88**	-0.54
Segment 4			**0.86**	0.63
Segment 5				**0.76**

Both sentences (n=90 contours).

	Segment 3	Segment 4	Segment 5	Segment 6
Segment 2	0.58	**-0.85**	-0.49	-0.17
Segment 3		-0.66	-0.52	-0.17
Segment 4			0.62	0.35
Segment 5				**0.75**

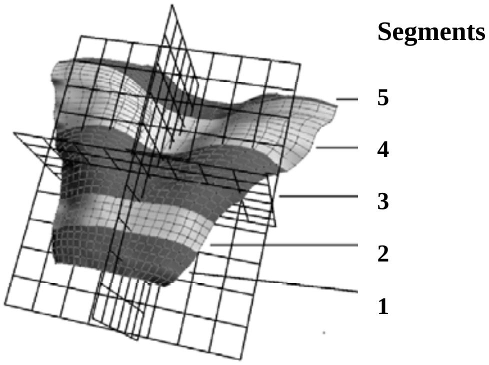


Figure 1. Coronal segments of the tongue. Adapted from Stone and Lundberg (1996).

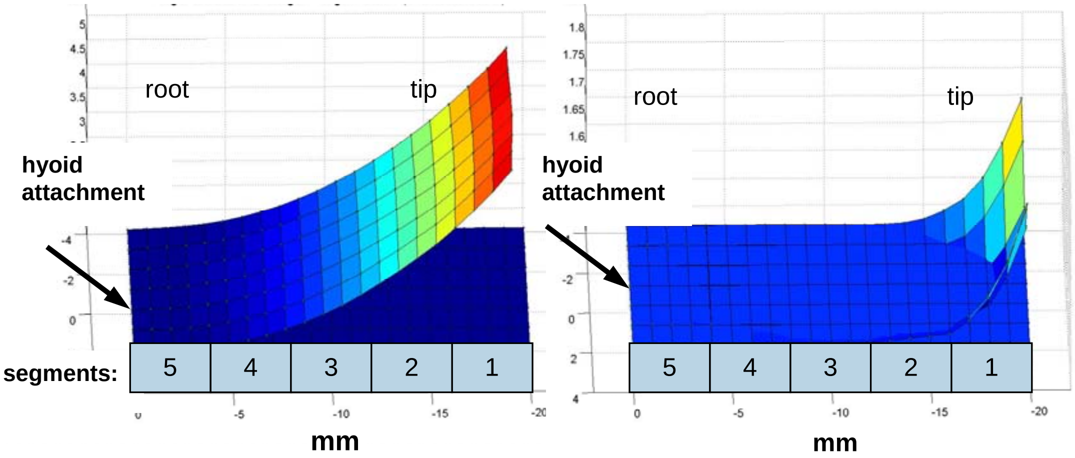


Figure 2. Retroflexion of the tongue tip created using different SL activations. A finite element model, which has 5 independently controlled tongue segments, begins with a flat surface. In both pictures GG and V are contracted at the tip (segment 1). On the left, SL is contracted in all five segments, on the right only in segment 1. The right is more consistent with retroflexion.

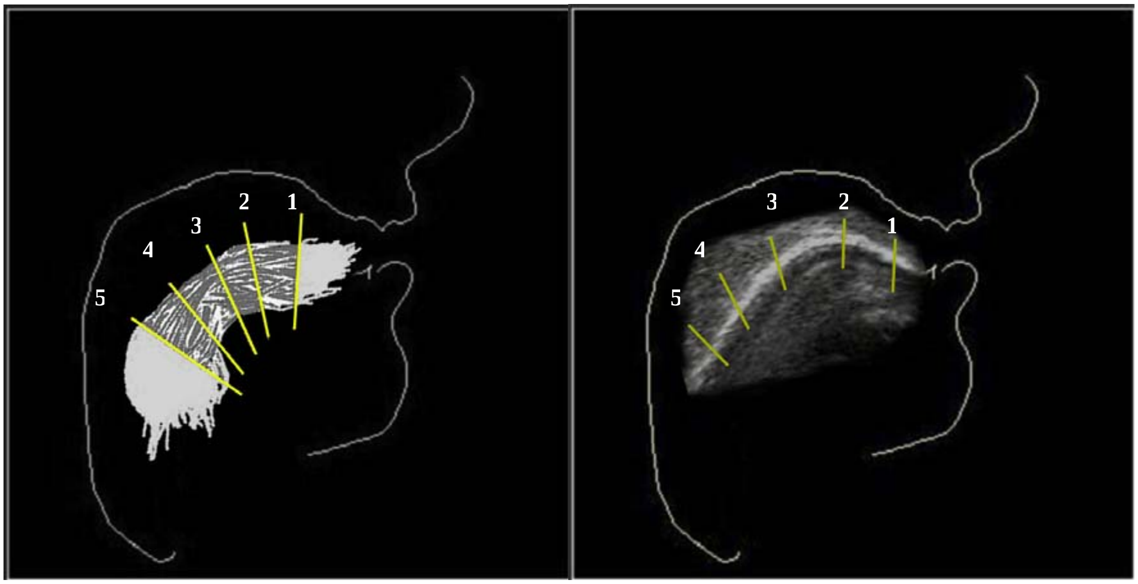


Figure 3. Five tongue segments for the Fisher-Logemann (left) and ran-rang (right) data sets.

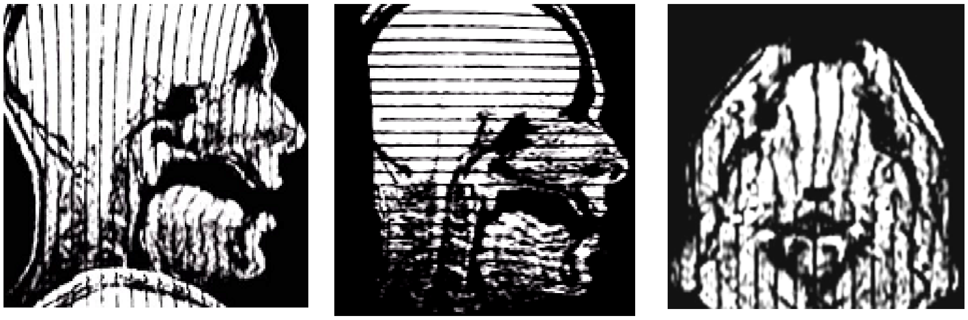


Figure 4. tMRI images of /ɑ/ with tags running (a) superior-inferior (a) medial-lateral and (b) anterior-posterior

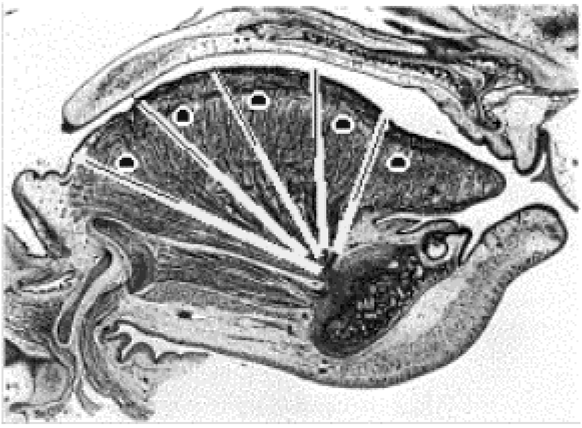


Figure 5. Locations of measurements for GG (heavy lines) and V (thin lines) and T (circles). Note that T is orthogonal to the picture.

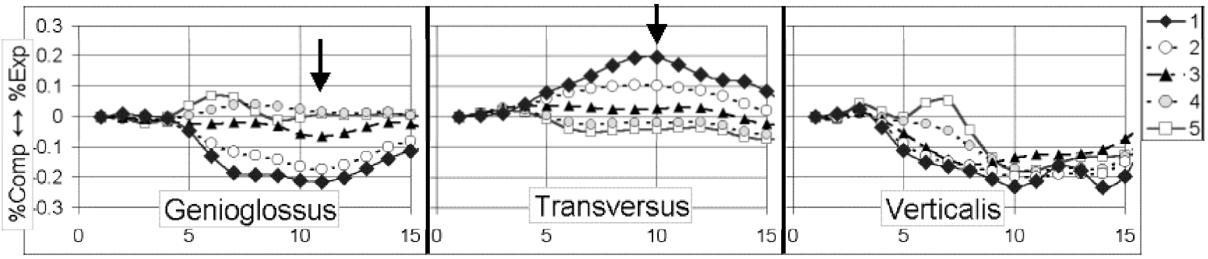
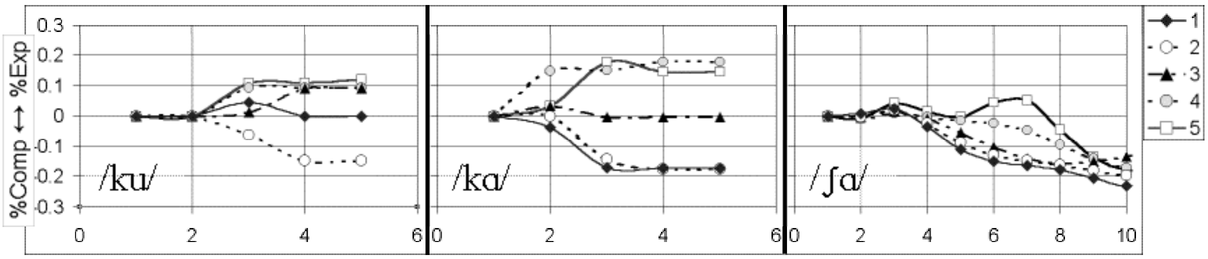
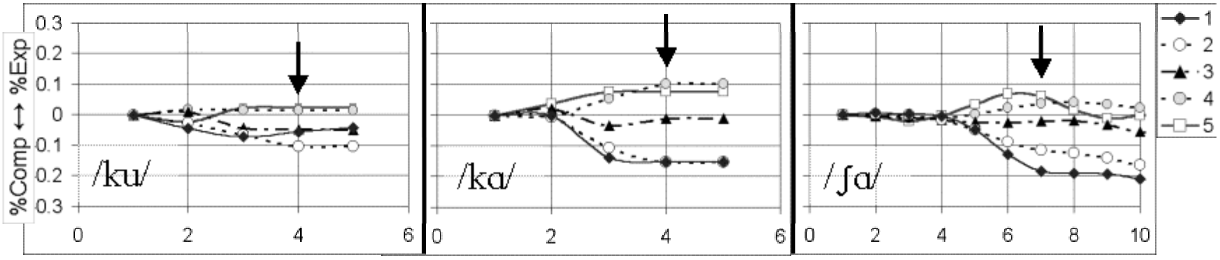


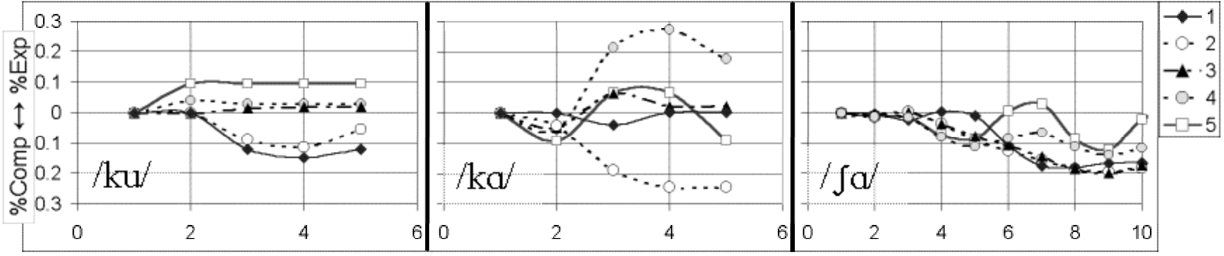
Figure 6. Strains in the line of action of GG, T and V-left for /ʃɑ/ spoken by subject 3. The most anterior segment, 1 (black diamonds), is most compressed in GG and V, and most expanded in T.



(a) Verticalis Left



(b) Genioglossus



(c)Verticalis Right

Figure 7. Compression over time for the three muscles Verticalis-left (top), Genioglossus (middle) and Verticalis-right (bottom) during the syllables (a) /kɑ/, (b) /ku/, (c) /ʃɑ/. N.B. /kɑ/ and /ku/ are spoken by subject 4, /ʃɑ/ is spoken by subject 3.