Taylor & Francis
healthsciences

Functional segments in tongue movement

MAUREEN STONE †, MELISSA A. EPSTEIN † and KHALIL ISKAROUS ‡

† Vocal Tract Visualization Lab, University of Maryland Dental School, Baltimore, Maryland, USA
‡ Haskins Lab, Yale University, New Haven, Connecticut, USA

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 (Ohman, 1967) postulated the tongue to be divided into tip and body executing quasi-independent motions. Geometric models of the tongue also assumed this division (cf.

Address correspondence to: Maureen Stone, Deparment of Biomedical Sciences, Rm, 5A12, University of Maryland Dental School, 666 W. Baltimore St., Baltimore, MD 21201, USA. E-mail: mstone@umaryland.edu
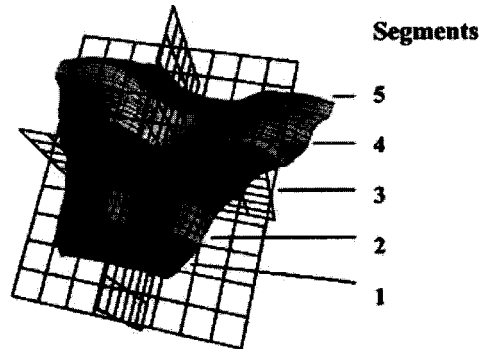
Figure 1.   *Coronal segments of the tongue. Adapted from Stone and Lundberg (1996).*

tongue differ from front-to-back in fibre 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 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 fibres. Rather it is a muscle of in-series design, in which short muscle fibres 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 fibres. If so, in-series fibres 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 upper portion of the tongue in an unbent state, i.e. shaped like an uncurved rectangular solid. The FEM has five segments that can be controlled independently (Essex-Torcasso 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.

The large number of motoneurons, the complex fibre orientation and fibre type,

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 five surface locations, for a phonetically balanced speech sample

The ultrasound study 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.

Experiment 1 tested the independence of tongue segments using five of the 14 sentences in The Fisher Logemann Test of Articulatory Competence. 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
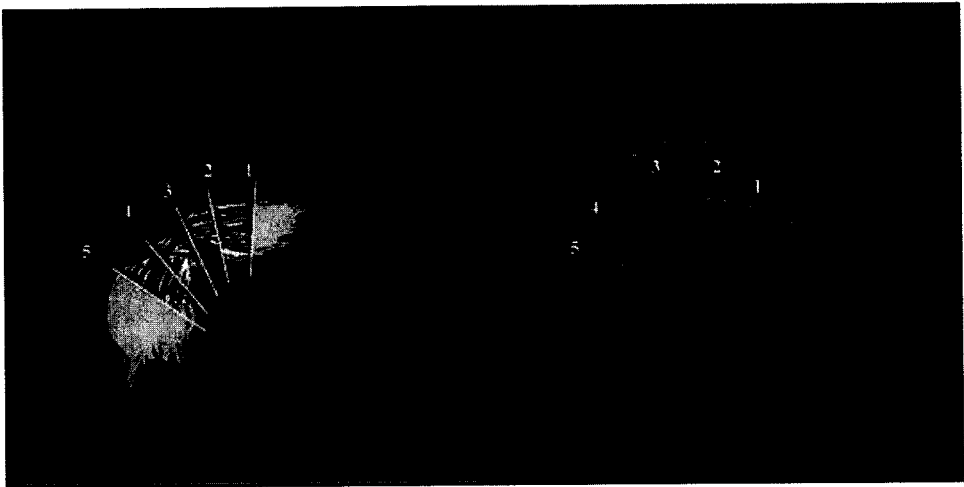

Figure 3.   *Five tongue segments for the Fisher-Logemann (left) and ran-rang (right) data sets.*

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**

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 (Harshman *et al.*, 1977; Sanguineti, Laboissiere and Ostry, 1998; Maeda, 1989). The data for factor analysis studies consist of either static tongue surface shapes (Harshman *et al.*, 1977; Maeda, 1990) 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.

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. 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 or horizontal fibres, with separate motoneurons. It is worthwhile, therefore, to examine differences in the local behaviour of these muscles.

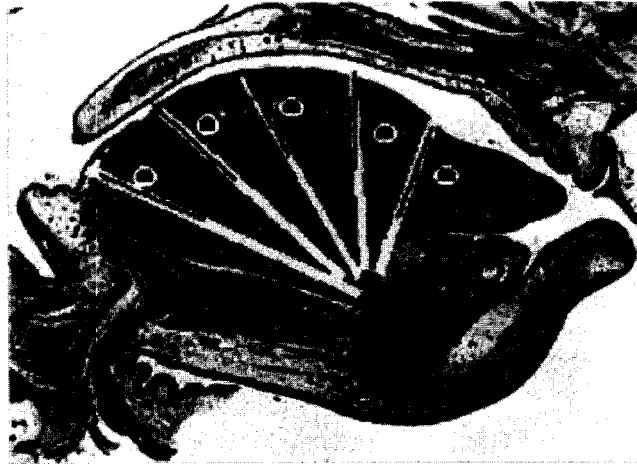It is known that GG has an architecture that could allow local activation at

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

were collected in three sagittal planes. This data set is discussed in detail in Stone, Davis, Douglas, NessAiver, Gullapalli, Levine and Lundberg (2001a, 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 segments 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 three muscles were measured according to their fibre 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.

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 at the maximum /ɑ/ position (arrows). The GG compression pattern was consistent with a single contraction occurring at the anterior-most segment (1) and propagating to neighbouring 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 4c).

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,
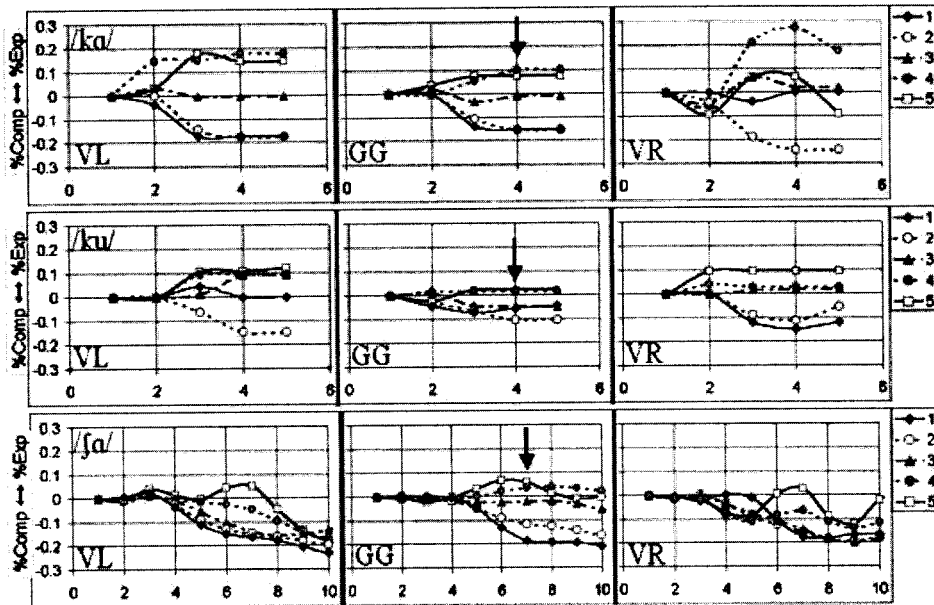
Figure 7.   *Compression over time for the three muscles Verticalis-left (VL), Genioglossus (GG) and Verticalis-right (VR) during the three syllables /ka/ (top), /ku/ (middle), /ʃa/ (bottom). N.B. /ka/ and /ku/ are spoken by subject 4, /ʃa/ is spoken by subject 3.*

compressed, that is, coupled to create a high-back to low-back motion. For /ʃa/ 1 was more compressed than 2, creating a high-front to low-back motion. This also may reflect a subject effect since /ʃa/ was spoken by subject 3, and /ka/ and /ku/ by subject 4.

Recall that for subject 3, V showed unitary contraction from /ʃ/ to /a/. For subject 4, V differed left-to-right and front-to-back. Left-to-right asymmetry reflected tongue rotation. In both /ka/ 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 /ka/, 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 behaviour for V and T, similar to GG (subject 4) in that their anterior and posterior segments moved in opposition, and local segmental behaviour 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

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 consonants based on MRI and EPG data. Part 1: The laterals. *Journal of the Acoustical Society of America*, **101**, 1064–1077.

NOORANI, S. M., 2003, Effect of Simulated Bite Opening on Swallowing Patterns in Normal Adults. MA thesis, University of Maryland Dental School.

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**, 627–634.

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., SMITH, P., STONE, M. 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. 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., 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., 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.

TAKEMOTO, H., 2001, Morphological analyses of the human tongue musculature for three-dimensional modelling. *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 and J. Bransford (Eds), *Perceiving, Acting and Knowing: Toward an Ecological Psychology* (Hillsdale, NJ: Lawrence Erlbaum Associates), pp. 211–265.

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.

WOZNIAK, W. and YOUNG, P. A., 1969, Further observations on human hypoglossal nerve. *Anatomischer Anzeiger*, **125**, 203–205.

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.