ISCA Archive
http://www.isca-speech.org/archive

3rd International Conference on
Spoken Language Processing (ICSLP 94)
Yokohama, Japan
September 18-22, 1994

TONGUE-PALATE INTERACTIONS IN CONSONANTS VS. VOWELS

Maureen Stone* and Andrew Lundberg**

*Dept of Electrical and Computer Engineering, *Department of Cognitive Science,
**Department of Computer Science, The Johns Hopkins University, Baltimore,
MD, 21218, USA.

Consonants and vowels have often been described as two parallel systems of production. To make vowels, the articulators produce slow or tonic movements. To make consonants, the articulators use fast or phasic movements.

This paper examined differences in 3D tongue surface shapes for consonants vs. vowels. 3D tongue surface shapes were reconstructed from 20 cross-sectional slices of the tongue surface, each 3° apart in a radial array. Steady state consonants and vowels were measured. The data supported the theory that for vowels, the tongue behaves like a muscular hydrostat. Local expansion and compression patterns follow tongue muscle morphology, and tongue shape is predictable from tongue location. However, for consonants, simple volume preserving deformations do not adequately explain the shape changes. Location, shape and pressure of tongue-palate contact must be known in order to adequately represent tongue shape for consonants.

Consonants and vowels have been hypothesized to derive from two parallel and different systems of production (cf. Ohman, 1966). Support for the different systems idea comes from differences in consonant and vowel production. For example, vowels are syllable nuclei with an open vocal tract, large airflows and small intra-oral pressure. Consonants are usually syllable margins with obstructed vocal tracts, small airflows and large intra-oral pressures. Vowel durations are long and affected by rhythm and rate. Consonants are short and unaffected. Vowels have low frequency spectral energy and one sound source. Consonants have high frequency spectral energy and up to three sound sources. The fact that these two "systems" are produced in alternation leads one to consider what advantage that might create in the production of speech. The present paper considers the role of tongue shape in vowels vs consonants, and the physiological, aerodynamic, and acoustic consequences of these roles. In order to fully capture consonant-vowel tongue shape differences, the shape of the entire 3D tongue surface must be considered. Therefore 3D reconstructions of tongue surfaces were created from ultrasound images of the tongue.

Reconstruction of 3D tongue geometry. Ultrasound images of the tongue surface were collected for static speech sounds using a developmental 3D ultrasound machine. The 3D ultrasound transducer has a single array of 128 ultrasound crystals mounted on a motorized pivot, allowing it to sweep a 3D space. The transducer collects 60 ultrasound slices, each 1° apart, in about 10 seconds. An ultrasound image is a visual representation of density changes in the 2D slice along the transducer crystal array axis. The surface of the tongue, being a tissue-air interface, is the most dramatic density change in the scan, and is visible as the lower surface of a bright line.

A custom edge detection program detected the surface contour of the tongue in each of the slices. These surface contours were stored as a series of 2D points in the plane of the slice. The detected surface points in each slice were then restored to their relative 3D coordinate locations. Since the geometry of the 3D transducer is known, we were able to locate its virtual pivot with relation to the slices. Twenty slices, each 3° apart, were rotated the appropriate number of degrees and then restored to their 3D positions.

Once the 3D points were aligned, we derived the simplest smooth surface on which they lay. Since the tongue is smooth and continuous, bsplines were used to fill the curves between the points. An interpolating bspline will pass through all the points in the slice with the simplest smooth curve. We did this for each slice, and also across all the slices, to obtain a grid of splines. The grid was then filled out with additional control points to obtain a bicubic piecewise bezier patch. The bezier patch describes the continuous tongue surface, and can be probed for statistical analysis, and visualized for more intuitive presentation. Examples of the resulting surfaces are shown in Figures 1 and 4 below.

Vowel Shapes. Vowels, particularly non-high vowels, use little or no tongue-palate contact. Their tongue shapes are therefore almost exclusively the result of muscular contraction. In such cases the tongue behaves like the muscular hydrostat model proposed by Smith and Kier (1989). That model is based on the behavior of hydrostatic animals, such as squid or octopus, which use fluid filled sacks as skeletons. The tongue has no sack, but rather uses its internal musculature to position the tongue, and to act as a skeleton thus stiffening and shaping the tongue. Hydrostats are independent, individual entities that do not need the additional support of rigid structures to facilitate their activity.

During non-high vowels the tongue is easily represented as a muscular hydrostat. The tongue is essentially composed of fluid-like muscles that are incompressible and thus volume preserving. Therefore, tongue shape derives from the mechanical principals of a volume preserving fluid structure, namely, expansion in one location requires compression in another. For example, to produce /i/, the anterior tongue is thrust upward by contracting Genioglossus Posterior muscle. That contraction also causes a deep posterior groove. In other words the anterior expansion requires posterior compression (see Figure 1).

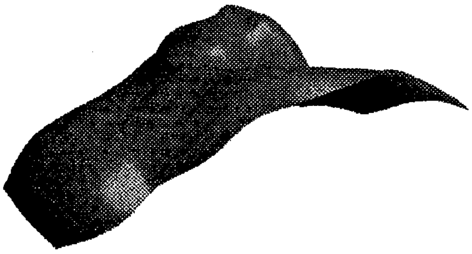

Figure 1. Three-dimensional reconstruction of [i] based on 20 ultrasound slices of the tongue each 3 degrees apart. Anterior is on the left.

Not only are tongue shapes during vowels related to tongue locations, but changes in tongue shape create tongue movements. Shape deformation is the only way a free-standing, incompressible, volume preserving structure like the tongue can be self-propelled. Accordingly, tongue shape and position during vowels vary systematically; a higher tongue has a steeper posterior slope (/i/ vs. /ɑ/), a more anterior tongue body uses a more anterior tongue root (/i/ vs /u/) root, and at the most displaced point on the tongue, cross-sectional shape becomes more concave with increased narrowing of the constriction (see Figure 2).

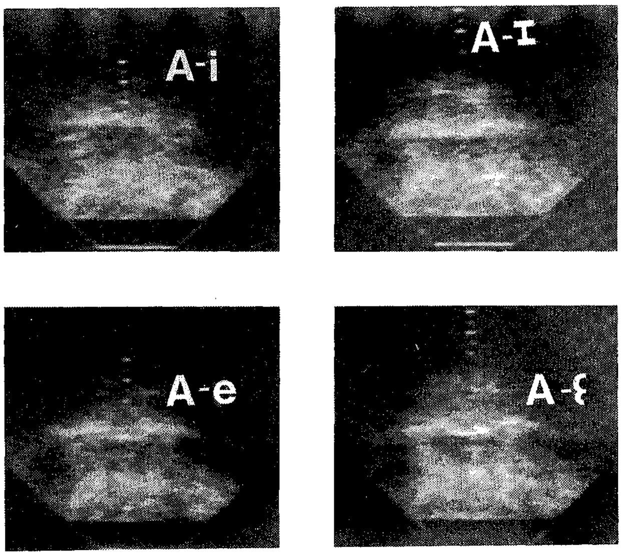

Figure 2. Cross-sectional images of the anterior tongue during [i], [ɪ], [e], [ɛ]. The lowest tongue contour [ɛ] has the deepest groove. (from Stone, et al., 1988, with permission).

Furthermore, the acoustic consequences of these changes are predictable; a higher tongue has a lower F1, a more anterior tongue has a higher F2, etc. The result is that models of tongue behavior during vowels abound (Hashimoto and Sasaki, 1982, Maeda, 1990, Mermelstein, 1973, Perkell, 1974), because knowing how the point vowels behave, allows us to predict quite accurately how the intermediate vowels will behave. Moreover, the more similar in tongue location two vowels are, the more similar will be their shape and spectra.

Consonant Shapes. Consonants do not show a systematic location-to-shape or location-to-acoustic relationship. For example, many consonants use a lingua-alveolar constriction, such as /t,l,s,n/. However, the commonality of constriction location does not guarantee similarity of tongue shape (see Figure 3), constriction shape, or acoustic spectrum.

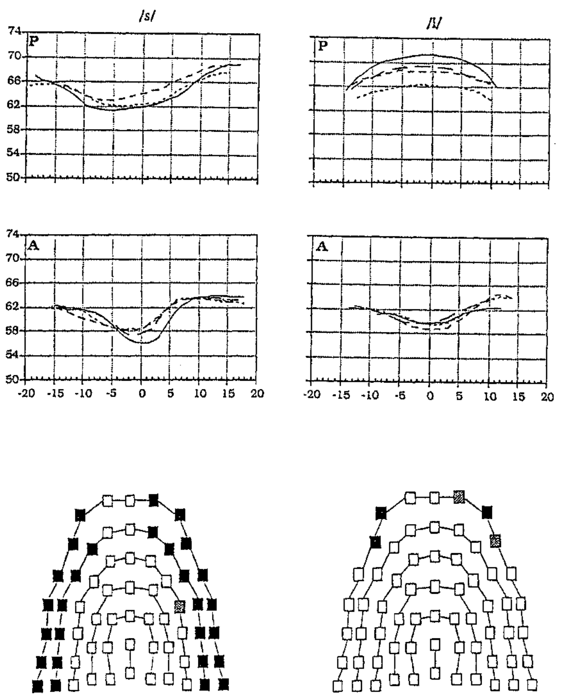

Figure 3. Tongue-palate contact patterns and two cross-sectional slices of the tongue for [s], [l]. (From Stone, et al., 1992).

For example, when comparing /l/ with /s/, the posterior tongue shapes are quite different. Moreover, cross-sectional shape reflects direction of airflow rather than constriction narrowness. Of course, the resultant spectra for these two sounds are quite different as well. Thus, palatal contact patterns directly affect 3D tongue surface shape, local cross-sectional shape, and the acoustic spectrum (cf. Stone et al, 1992).

In the case of consonants the tongue does not behave solely as a muscular hydrostat; it does not control all its behaviors using its muscles. The tongue also uses the resistance afforded by the palate to fine tune both its own shape and that of the vocal tract. 3D tongue surface shapes for consonants do not simply reflect volume preserving principles, but also the effects of tongue/palate contact. Figure 4 shows a midline groove created to direct a narrow jet of air onto the teeth. Genioglossus contracts throughout the tongue. This, in conjunction with complete lateral tongue-palate closure causes the midline air channel. Genioglossus Anterior must contract more forcefully than for [i] in order to create a channel with just the right degree and shape, in the more anterior, lower, alveolar part of the palate (see Figure 4).

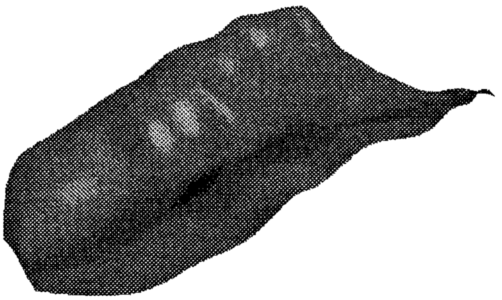

Figure 4. Three-dimensional reconstruction of [s]. Anterior is on the left.

The use of local tongue-palate contact to shape the tongue and create spectral disturbances during consonants, but not vowels, is consistent with other evidence that consonants and vowels are different subsystems (cf Ohman, 1966).

High vowels. For high vowels like /i/ and /u/, tongue-palate contact, although as great as during many consonants, does not appear to increase the complexity of tongue shape, neither does it create sound or divert airflow. Tongue-palate contact patterns among vowels are different from each other in magnitude, more than shape (Recasens, 1991). Models have been able to create an accurate tongue shape and position for /i/ using tongue muscular contraction alone (Fujimura and Kakita, 1979, Perkell, 1974). In that case the palate appears to be simply a limiting factor, providing an upper boundary for the tongue just as the teeth and floor of the mouth provide lateral and lower boundaries for other vowels. These boundaries provide both outer limits and proprioceptive feedback, useful more for accurate tongue positioning, and less for shape manipulation.

Support for two different production systems. A gross physiological distinction between consonants and vowels would be to consider the tongue to be **fixed** at one or more points during consonants, and quasi **free**-floating during vowels. The consequences of fixed vs. free postures are as follows. Physiological differences. When fixed, tongue-palate bracing can increase tongue shape complexity, but when free, shape and location are dependent. When fixed, coarticulation occurs at unattached regions primarily, but when free, coarticulation occurs globally. Aerodynamic Differences. When fixed, the tongue constriction may redirect airflow, but when free, the constriction seldom redirects airflow. Acoustic Differences. When fixed, the constriction may act as a sound source, but when free, the constriction produces no sound. Possible Effects of these Differences. (1) Faster transitions between adjacent phones. (2) Less fatigue when alternating behavior patterns.

Conclusions

The tongue produces shapes differently for vowels and consonants. For vowels the tongue behaves as a volume preserving muscular hydrostat. Local expansion and compression patterns follow patterns of tongue muscle morphology, and tongue shape is predictable from tongue location. For consonants, hydrostatic behavior combines with the bracing effects afforded by tongue-palate contact. Local expansion and compression patterns are affected by location and force of palatal contact, not just tongue morphology. Tongue shape is not predictable only from constriction location, palatal contact pattern must also be known.

References

Fujimura, O., and Kakita, Y. (1979) Remarks on quantitative description of the lingual articulation. in Frontiers of Speech Communication Research., ed by B.Lindblom, and S. Ohman, (London: Academic Press) p. 17-24.

Hashimoto, K., and Sasaki, K. (1982) On the relationship between the shape and position of the tongue for vowels. J. Phonetics, 10, 291-299.

Maeda, S. (1990) Compensatory articulation during speech; evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model. in Speech Production and Speech Modelling, ed by W. Hardcastle and A. Marchal, p.131-150.

Mermelstein, P. (1973) Articulatory model for the study of speech production. J. Acoust. Soc. Amer., 53, 1070-1082.

Ohman, S. (1966) Coarticulation in VCV utterances: Spectrographic measurements. J. Acoust. Soc. Amer., 39, 151-168.

Perkell, J. (1974) A physiologically-oriented model of tongue activity in speech production. Ph.D. Dissertation, MIT.

Recasens, D. (1991) An electropalatographic and acoustic study of consonant-to-vowel coarticulation. J. Phonetics, 19, 177-192.

Smith, K. and Kier, W. (1989) Tongues, trunks and tentacles: moving with skeletons of muscle. American Scientist, 77, 29-35.

Stone, M., Talbot, T., Shawker, T., and Rich, A. (1988) Cross-sectional tongue shape during the production of vowels. J. Acoust. Soc. Amer., 83, 1586-96.

Stone, M., Faber, A., Raphael, L., and Shawker, T. (1992) Cross-sectional tongue shape and lingua-palatal contact patterns in [s], [ʃ], and [l]. J. Phonetics, 20, 253-270.