Palate shape effects on characteristic vowel tongue postures

Mark K. Tiede

ATR Human Information Processing Labs., 2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0288 Japan

Abstract: In this work MRI vocal tract data obtained during sustained vowel production is used to investigate the interaction between individual speaker differences in hard palate shape and the compensatory adjustments to tongue posture needed to tune language-specific vowel formant frequencies. PCA techniques were used to characterize 3D tongue and palate shapes for vowel inventories of Japanese subjects. Results show that the axis of tongue body displacement is associated with oral cavity depth (OCD): shorter OCD subjects showed greater vertical /a/ vs. /u/ differences, and longer OCD subjects used more of a horizontal displacement component. These results are consistent with the speaker-specific articulator ('A' space) adaptations suggested in previous work (Honda et al. 1996).

MOTIVATION

Human speech articulation necessarily occurs within bounds imposed by anatomical structure, and languages exploit aspects of that structure to produce characteristic and contrastive sounds. Because vowel sounds are distinguished by harmonic resonances tuned by the posture of the tongue along the length of the supraglottal vocal tract, individual differences in hard palate shape must be matched with compensatory adjustments to tongue posture to preserve language-specific vowel formant patterns. In this work 3D morphological data obtained from Magnetic Resonance Imaging (MRI) are used to examine such compensatory interaction.

A-SPACE

The starting point used here for cross-subject comparison is the Articulator or 'A' space introduced by Honda et al. (1). The A-space is a trapezoid bounded by the palatal plane, the rear pharyngeal wall, and the base of the mandibular symphysis that can be used as a framework for scaling the articulatory gestures of individual speakers. As used here the length of its upper side is determined by the line from the maxilla to the basion intersecting the rear pharyngeal wall, giving an index of Oral Cavity Depth (OCD). The distance of a perpendicular from this line to the menton (the lowest point on the mandibular symphysis) gives an index of lower facial height (LFH).

Honda et al. used measurements obtained from X-ray microbeam raster scans of 10 Japanese and 10 American English speakers to show a weak negative correlation between OCD and LFH. Their evidence also suggests that speaker OCD is related to differences in vowel articulation: subjects with shorter OCD showed a pattern of greater vertical tongue difference, while longer OCD subjects showed greater horizontal difference.

In the current study similar measurements were taken from midsagittal MR images obtained for 12 Japanese (including 2 females) and 12 Caucasian subjects (including 3 females), in a closed-mouth (nonphonating) condition. These were augmented by an additional metric, an index of palatal width (PW) obtained from

FIGURE 1. A-space derivation

FIGURE 2. Regression of PW scaled LFH x OCD
transverse MR images of the same subjects, measured as the distance between the interior edges of the first true molars on both sides of the upper jaw.

Regression of the OCD against LFH values obtained gave a weak positive correlation ($r^2 = .379^* )$ due mostly to the Caucasian subgroup ($r^2 = .649^*$; Japanese subgroup n.s.). Scaling each of these measures by the associated PW improved both overall correlation ($r^2 = .616^*$), and the Caucasian ($r^2 = .740^*$) and Japanese ($r^2 = .610^*$) subgroup correlations (see Figure 2). This pattern of results suggests that while larger A-spaces are in general associated with larger values along each dimension, disproportionately large values for one dimension tend to be offset by relatively smaller values along one or both of the other space parameters.

**VOLUME ANALYSIS**

Previous work (2) has established a PCA methodology permitting concise characterization of 3D vocal tract shape derived from MRI volumes. With this approach cross-sectional shapes are extracted from volume data by sampling along a semipolar grid system used to 'unwrap' the tract. Using SVD, each cross-section is approximated by a weighted sum of two eigenvectors. Since adjacent tract sections are correlated, section weights can themselves be approximated by a weighted sum of four eigenvectors.

Here this approach has been applied to a subset of the A-space subjects for whom sustained vowel data were available: 4 Japanese subjects (3 male, 1 female), producing the five Japanese vowels /a/, /i/, /u/, /e/, /o/. Four principal components sufficed to account for 83% of the total variance of the combined data corpus, with the first component alone accounting for 62.6%.

![Figure 2](image.png)

**FIGURE 2.** Midsagittal and cross-sectional projections of 1st component variance (+/− 1 std. dev.)

The figure shows midsagittal and cross-sectional projections of the 1st component variance (+/− 1 s.d.) for the subjects with longest and shortest measured OCD values. A comparison of the centroids of derived tongue position shows that subjects with shorter oral cavity depth use more of a vertical displacement component. The results also suggest that subjects with wider palatal width available do not need to displace the tongue body to the same extent as those subjects with narrower palatal cross-dimensions.

**REFERENCES**


1286