EUROSPEECH '97
5th European Conference on Speech
Communication and Technology
Rhodes, Greece, September 22-25, 1997

NEW RESULTS IN VOWEL PRODUCTION: MRI, EPG, AND ACOUSTIC DATA

Shrikanth Narayanan[1], *Abeer Alwan*[2], *Yong Song*[2,*]

[1]AT&T Labs–Research, 180 Park Avenue, Florham Park, NJ 07932
[2]Department of Electrical Engineering, UCLA, 405 Hilgard Avenue, Los Angeles, CA 90095

ABSTRACT

MRI, EPG, and acoustic data for the vowels /a, i, u/ are analyzed. The vocal tract geometry, tongue shapes, and inter-subject variability are studied. The data are used for studying the articulatory-acoustic relations for these sounds.

1. INTRODUCTION

One of the main challenges in speech production modeling is obtaining physiological (or articulatory) data that are accurate and reliable for both qualitative and quantitative analyses. Examples of such data include vocal-tract dimensions, tongue shapes, and position and velocity of the lips, jaw and tongue body during speech production. In this paper, we report new results and models for vowel production based on the articulatory information obtained using techniques such as magnetic resonance imaging (MRI) and dynamic electropalatography (EPG) in conjunction with acoustic data. The data are used for studying articulatory-acoustic relations.

MRI, EPG, and acoustic data were collected for the three-point vowels /a, i, u/ from four subjects. Although this study focussed only on /a, i, u/, it complements, and offers many advantages when compared to, other recent studies on vowels [1, 2]. Our study included both male and female subjects, and had extensive MRI data (acquired in all 3 anatomical planes). The MRI data reported in this paper provides further insights on 3D tongue shapes for vowels investigated using ultrasound imaging [3]. EPG data were also included to enable intra- and inter-speaker variability analyses. MRI scans require artificial prolongation of the sound but they provide information on the shape of the vocal tract not obtainable by other methods. Electropalatography provides dynamic articulatory information but is restricted to linguopalatal contact from the alveolar region to the anterior part of the velum. Nevertheless, together the articulatory data available from these two techniques, along with the acoustic data, enable us to obtain an increased understanding of vocal tract and tongue shape mechanisms and articulatory-acoustic relations of vowels.

Data were also collected with the subjects in both supine and upright positions (in order to verify the effects of supine position used during MRI experiments). In addition to providing 3D vocal tract renditions and length/area/volume measurements (including those of piriform sinuses), our analysis includes 3D tongue shape description. The articulatory data are then used in acoustic modeling. These vowel data form a part of a large data corpus of an ongoing articulatory-acoustic study series that also includes consonants such as fricatives [4], laterals [5], and rhotics [6] from the same subjects.

*Now with PairGain Technologies, Los Angeles, CA

2. METHODOLOGY

Four phonetically-trained, native American English speakers [2 males and 2 females] served as subjects.

MRI: Information about the 'static' vocal-tract shapes came from MRI scans (GE 1.5T scanner) at contiguous 3 mm intervals in the sagittal, coronal, and axial anatomical planes, which allowed three dimensional views of the vocal tract and tongue shapes to be constructed in a computer representation. Measurements of vocal tract length, area functions, and cavity volumes were also obtained. Details of image acquisition and analysis are similar to those given in [4].

EPG: EPG data were recorded using *Kay Elemetrics Palatometer*. Each subject has a custom-fitted acrylic palate with 96 sensing electrodes. The speech material consisted of the vowels /a, i, u/, produced in sustained utterances, similar to the MRI experiments. Data were recorded with subjects in both supine and upright postures while speaking. Eight repetitions of each condition were obtained.

3. DATA ANALYSIS

The analysis of the raw MRI and EPG data is presented first followed by a discussion on 3D vocal tract and tongue shapes.

MRI data: Midsagittal MR images for /a, i, u/ (Fig. 1(1)) enable a detailed comparison across the different vowels and speakers. In addition to showing the well-known tongue body shape for these sounds (low-back for /a/, high-front for /i/ and high-back for /u/), these images clearly show the tongue root and the epiglottal positions: Retracted tongue root with decreased epiglottal vallecular volume in /a/ while advanced tongue root with a distinct epiglottal space in /i/. The tongue root behavior for /u/, a high-back vowel, is similar to, but is less advanced than, that of /i/. For high vowels, a more anterior tongue body is associated with a more advanced tongue root, an observation consistent with previous studies. However, such relations in consonant sounds with similar high anterior tongue body are not straightforward: Some consonants show behavior similar to vowels while in others, it seems more complex. For example, while both /ʃ/ and and /ʒ/ have anterior tongue body behavior similar to /i/, only the voiced sound /ʒ/ shows prominent tongue root advancement [4] paralleling /i/. Similarly, in the production of American English /r/s, which also show anterior tongue body behavior similar to /i/, both an interplay between the location of the constriction made by the tongue body and the degree of 'pharyngealization' and a tendency toward a systematic manipulation of the anterior vs. posterior tongue surface concavity were observed [6]. The prominent presence of the epiglottis, such as seen in /i/, may be attributed to cause turbulence generation in the pharyngeal cavity (aspiration) during speech production [7].

The coronal and axial cross-sections reveal the tongue shape and the vocal tract cavities in much greater detail. Sample coronal cross-sections for /a/ (MI) are shown in Fig. 2. Computer recon-

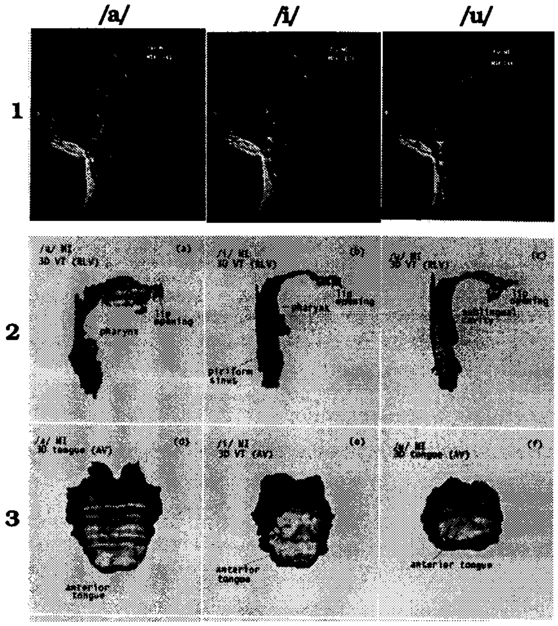

Figure 1: *MRI data for vowels /a,i,u/ of a male subject (MI). (1) Mid-sagittal images (2) 3D vocal tract images, lip opening toward the right of each panel (3) Anterior view of 3D tongue shapes, tongue tip toward the bottom of each panel.*

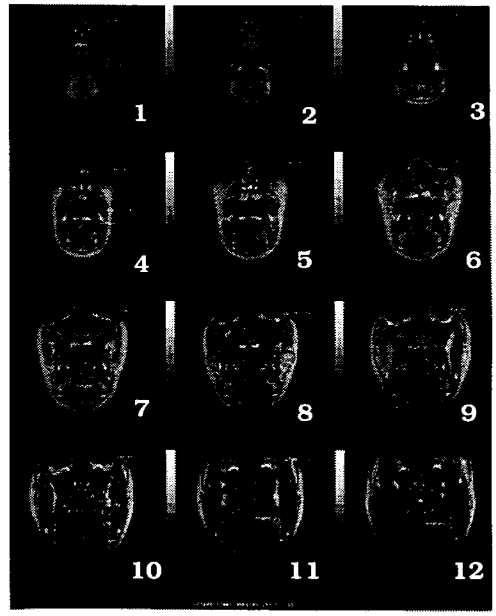

Figure 2: *Coronal MRI images of the oral cavity for /a/ (MI). Panel 1 is a cross-section at the lips while panel 12 is at the velar region. Note the gentle tongue surface grooving along the mid-sagittal line through most of the middle tongue body (5-10). The lateral bracing of the tongue against the roof of the oral cavity can be seen in panels 11-12.*

struction that simultaneously linked raw scans from the axial, coronal, and the sagittal planes enabled obtaining oblique sections along any arbitrary direction. This allowed for interactively studying the cross-sectional vocal tract and tongue shapes, and obtaining accurate measurement of the area functions. Coronal slices clearly show the oral cavity cross-sections: large areas in /a/ compared to small areas in /i/ due to the low-front/high-front contrast in the tongue body position. The tongue body surface in /a/ might show a slight grooving along the mid-sagittal line, a tendency which appears to be subject dependent [subjects MI (Fig. 2(6-9) and Fig. 1(3)) and SC (not shown here)]. For /u/, there is an increased front cavity volume in some subjects due to a sublingual space (Fig. 1(1)). The anterior region for /i/ and /u/ has similar characteristics for all subjects while /a/ is most variable in that region (for example, Fig. 3). Constriction lengths and areas for /i/ are comparable across subjects (l_c = 3.3-3.5cm, A_c = 0.3-0.4cm^2). The maximum area of the front cavity for /a/ is, depending on the subject, between 4.0-6.8cm^2 and for /u/ it is 2.9-6.1cm^2. Lip opening is found to be speaker-dependent.

The axial scans show the details of the pharyngeal and the laryngeal regions. The presence of special tissue structures such as the uvula, epiglottis, and the aryepiglottic folds contribute to creating cavity structures along different regions of the pharyngeal airway. The presence of the uvula decreases the effective cross-sectional area around the vocal-tract bend. The presence of epiglottis, prominent in /i/ and /u/ due to tongue root advancement, creates a distinct vallecula separated entirely from the pharyngeal airway. The aryepiglottic folds create the piriform sinuses on either side of the laryngeal vestibule.

The inter-subject differences apparent in the front cavity areas of /a/ when compared to /i/ (Fig. 3 a,b) are attributed to anatomical differences across subjects which play a prominent role due to the low tongue body position in /a/ when compared to the high-front tongue body making the front approximation in /i/. On the other hand, there is less inter-subject variability in the pharyngeal areas of /a/ when compared to /i/. The inter-subject area function variability in /u/ is much higher than /a/ and /i/: the variability in the front cavity is similar to /a/ and in the back cavity, somewhat similar to /i/. These observations suggest that there is an increased apparent inter-subject variability in area function values in vocal tract regions whenever there is movement of articulators away from one another since individual anatomical differences are allowed to show prominently. The reverse is the case when articulators move toward each other.

EPG data: Analysis of the linguopalatal contact patterns showed lingual contact may appear anywhere in the hard palatal region (note that the pseudopalate used in EPG recordings extends only until the beginning of the velum). The percentage of total electrodes contacted during the production of the sustained vowels is used as the measure for comparative analyses.

There was little (subjects MI, SC, AK; maximum observed contact 12%) or no linguopalatal contact (subject PK) during the production of /a/ (Fig. 4(a)). The contact was mainly in the back palatal region, and the EPG data confirms the lateral bracing of the posterior tongue body against the roof of the oral cavity seen in the MRI data for /a/ of some subjects. There is some effect of the supine posture while speaking observed in the (back palatal) linguopalatal contact patterns although it was found to be a subject-dependent behavior (subjects MI, SC). These data indicate that subjects who reveal posterior tongue body raising and a concomitant lateral tongue bracing in the back palatal region (subject-dependent posterior tongue raising and/or tongue root backing that is accompanied by lateral tongue bracing has also been observed in consonants such as dark /l/ and /r/ in American English [5, 6]), tend to a slightly exaggerate this bracing as a way of compensating while speaking from a supine position. This perhaps contributes to slight decrease in the corresponding area function values when compared to those obtainable in upright positions.

On average, about 50-70% of the total electrodes were contacted during the production of /i/ (Fig. 4(b)). Subject MI showed slightly more contact in /i/ produced while speaking from an upright posture while the reverse was the case for PK and SC. Subject AK showed no such posture effects. The supine/upright differences were, however, statistically insignificant.

Inter-subject differences in the actual region of linguopalatal contact were much higher in /u/ (Fig. 4(c)) than in /a/ and /i/, similar to the observations made earlier in the discussion of area functions: the contact regions were more toward the front in AK and SC when compared to those of PK and MI. On average 12-35% of the total electrodes were contacted with a slight tendency of more contact in SC and PK.

The symmetry in the contact patterns was largely subject dependent, and to a small extent sound dependent, and was not dependent on the position assumed while speaking. Asymmetry observed in our subjects in general showed more left-favored contacts, and the effects were, relatively, most prominent in /u/ and least prominent in /a/. Subjects MI and SC who showed posture effects in the lateral back-palatal tongue bracing, seemed to do so more on the left than on the right side.

3D shape analysis: The results of Stone et al [3] show that the tongue shaping (groove) characteristics of vowels are not affected by context but rather depend on vowel quality. A study of artificially sustained sounds, hence, could provide a fair idea of the 3D vocal tract and tongue shapes.

ANTERIOR TONGUE BODY: Our MRI data indicate that the anterior tongue body for /a/ tends to be somewhat concave (gentle grooving along the midsagittal line when viewed from above, from the palate), and the degree of concavity is subject dependent. Subject MI [Fig. 2(6-9) and Fig. 1(3)] showed distinct anterior tongue surface concavity along the midsagittal line while subject PK did not (Fig. 5(2)). Subjects with noticeable concavity tend to show some amount of lateral linguopalatal bracing, around the mid-tongue region (although the contacts are very small compared to those observed in consonants such as /s/ showing concave anterior tongue surfaces [4]).

The anterior tongue surface for /i/ and /u/ was convex ((Fig. 1(2), Fig. 5(2)), and similar across the subjects. Recall that these sounds were characterized by significant lateral linguopalatal bracing in the mid/back palatal region. In this respect (tongue shape and palatal bracing), the high vowels are similar to the sibilant fricative /ʃ/ rather than the previously speculated /s/. Although the lateral linguopalatal bracing does facilitate the creation of the above mentioned tongue surface shapes, that its role may not be necessary as it is for some consonant sounds.

POSTERIOR TONGUE BODY: In general, all vowels showed a non-convex posterior tongue surface. The extent of the concavity seems to be a subject-dependent behavior. The high-front vowel /i/ showed deeper posterior grooves than the high-back /u/ (Fig. 6). The observation that front vowels have deeper posterior grooves than the back vowels [3], however, may be a subject-dependent phenomenon: while subject PK's /i/ has greater concavity than her /a/ (Fig. 6(2)), the posterior tongue groove characteristics of subject MI's /a/ and /i/ were very similar (Fig. 6). The caveat for this subject is, especially if the groove characteristics of the front and back vowels were not dramatically different in the first place, that the compensatory linguopalatal bracing while recording the MRI data in the supine position could have contributed to the deeper groove in MI's /a/. Regardless, some subjects consistently show linguopalatal bracing in the posterior tongue region for /a/ even while in upright position of speaking and have prominent posterior tongue grooving (subjects SC, AK).

Similarity of tongue-palate contact patterns between sibilants and /i/ had previously led to the conclusion that high front vowels such as /i/ are grooved anteriorly like /s/, especially in /s/ contexts (for example, see discussion in [3]). Our data, however, show that /i/ and /u/ have anterior tongue body shapes and linguopalatal contact patterns similar to /sh/ and /r/, respectively. Nevertheless, it is not clear with the given data the extent of the necessity/importance of the tongue bracing against the palate toward creating the observed shapes.

		Male (MI)			Female (PK)		
	Formants	/a/	/i/	/u/	/a/	/i/	/u/
Natural speech	F1	650	260	390	950	340	390
	F2	1120	2240	1210	1550	2580	820
	F3	2370	2770	2120	2930	3530	2580
Maeda's model	F1	576	272	296	672	300	376
	F2	1040	1980	1208	1254	1984	1360
	F3	2290	2850	2088	2344	3272	2219
FTD model	F1	550	200	200	300	600	-
	F2	1190	990	1190	1370	2390	990
	F3	2490	2490	1790	3380	3180	1990

Table 1: Formant values in Hz from Maeda's models and finite time difference (FTD) models for /a,i,u/. Simulation of /u/ with explicit sublingual component in Maeda's model resulted in formants (F1, F2, F3) Hz: (284, 1168, 2076)Hz for MI and (230, 748, 2196)Hz for PK.

The 3D vocal tract shapes are shown in Figs. 1(1) and 5(1). The vowel /i/ has larger pharyngeal cavity volume when compared to /a/ due to retraction of the posterior tongue body. The anterior tongue body of /u/ is similar to that of /i/ in that it is convex but relatively farther away from the lips. The pharyngeal volume in /u/ is also smaller than in /i/. These vocal tract characteristics explain the area functions shown in Fig. 3. No gender-based differences in the vocal tract characteristics were found.

4. ACOUSTIC MODELING

Two different vocal tract simulations were used to synthesize the sustained vowels /a,i,u/ spoken by a male (MI) and a female speaker (PK) using the MRI-derived area functions (Fig. 3): Maeda's vocal tract simulation [8] and a simplified finite time difference (FTD) approach [9]. Sound wave propagation in both models reported in this paper are one-dimensional, although the FTD model in principle is not limited by planar wave propagation assumptions and can be expanded to 3D. A voicing function is applied to the glottis, and two different terminations were used at the lips: short circuit and radiating (modeled as a piston in an infinite baffle). Maeda's model includes frequency dependent losses while vocal tract losses were not included in the current FTD model.

Simulation results are provided in Table 1 along with measurements made from the subjects' natural vowel tokens. Maeda's model, in general, provides a fair agreement between the actual and predicted values of formants. In the case of of /u/, especially for subject PK, where an appreciable sublingual cavity was found, the explicit inclusion of MRI-derived sublingual areas (note sublingual cavity for /u/ in Figs.1 and 5) as a separate "side-cavity" in the simulation, yielding improved results for F2 and F3. Such a simulation of the sublingual cavity, however, resulted in undershoot of the predicted F1 values, indicating that this approximation is not entirely perfect.

The FTD model produced fair results in both frequency and time domain, suggesting a promising speech production modeling approach for further exploration. Nevertheless, results indicate that the 1D model is not adequate and a more detailed simulation employing the 3D information should be attempted in the future.

5. SUMMARY AND CONCLUSIONS

MRI and EPG data, including 3D vocal tract and tongue shapes, for the vowels /a,i,u/ were analyzed for 2 male and 2 female American English speakers. Inter-subject variability in the articulatory data were discussed. Finally, articulatory-acoustic relations were investigated using simulations of the vocal tract using Maeda's model and a finite time difference model.

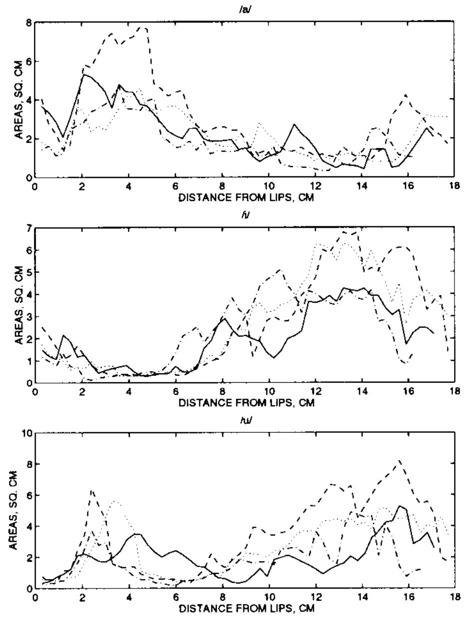

Figure 3: *Area functions for vowels /a, i, u/ for four subjects. Females: solid AK, dot-dashed PK; Males: dashed MI, dotted SC.*

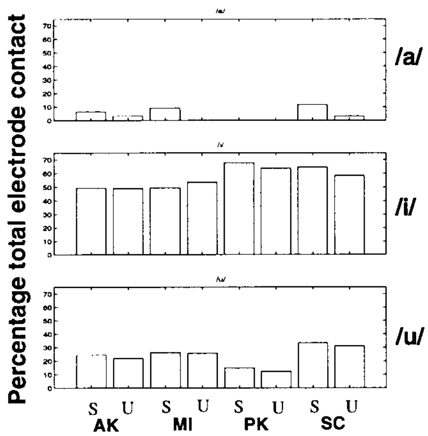

Figure 4: *Percentage total electrode contacted during the production of /a,i,u/ by the subjects AK, MI, PK, and SC in both supine (s) and upright (u) positions while speaking.*

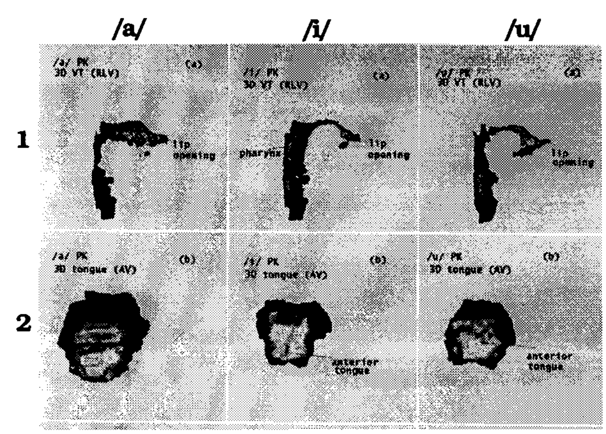

Figure 5: *MRI data for vowels /a,i,u/ of a female subject (PK). (1) 3D vocal tract images, lip opening toward the right of each panel (2) Anterior view of 3D tongue shapes, tongue tip toward the bottom of each panel.*

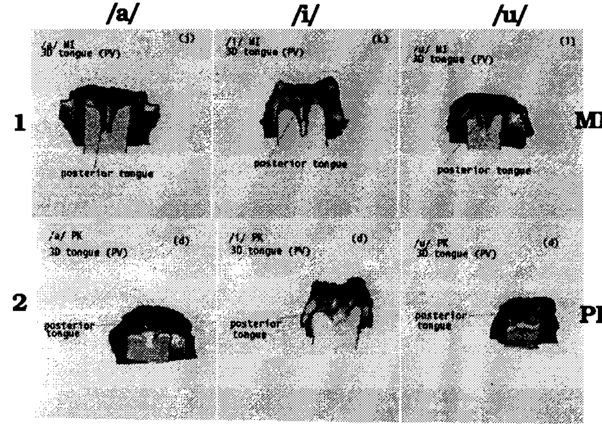

Figure 6: *Posterior view of the tongue body (looking from the pharyngeal wall) for vowels /a,i,u/. (1) Male subject (MI) (2) Female subject PK.*

Figure 7: Tube model for subject MI's /a/ with an extrapolated 1mm mesh thickness used in FTD simulations.

6. REFERENCES

[1] T. Baer, J. C. Gore, L. C. Gracco, and P. W. Nye, "Analysis of vocal tract shape and dimensions using magnetic resonance imaging: Vowels," *J. Acoust. Soc. Am.*, vol. 90, pp. 799–828, Aug. 1991.

[2] C. A. Moore, "The correspondence of vocal tract images with volumes obtained from magnetic resonance images," *J. of Speech and Hearing Research*, vol. 35, pp. 1009–1023, Oct. 1992.

[3] M. Stone, T. Shawker, T. Talbot, and A. Rich, "Cross-sectional tongue shape during the production of vowels," *J. Acoust. Soc. Am.*, vol. 83, pp. 1586–1596, 1988.

[4] S. Narayanan, A. Alwan, and K. Haker, "An articulatory study of fricative consonants using magnetic resonance imaging," *J. Acoust. Soc. Am.*, vol. 98, pp. 1325–1347, Sept. 1995.

[5] S. Narayanan, A. Alwan, and K. Haker, "Toward articulatory-acoustic models for liquid consonants based on MRI and EPG data. Part I: The laterals," *J. Acoust. Soc. Am.*, vol. 101, pp. 1064–1077, Feb 1997.

[6] A. Alwan, S. Narayanan, and K. Haker, "Toward articulatory-acoustic models for liquid consonants based on MRI and EPG data. Part II: The rhotics," *J. Acoust. Soc. Am.*, vol. 101, pp. 1078–1089, Feb 1997.

[7] C. H. Coker, M. H. Krane, B. Y. Reis, and R. A. Kubli, "Search for unexplored effects in speech production," in *Proceedings of ICSLP*, vol. 2, pp. 1121–1124, Oct. 1996.

[8] S. Maeda, "A digital simulation method of the vocal-tract system," *Speech Communication*, vol. 1, pp. 199–229, Oct. 1982.

[9] Y. H. Song, "Finite time difference simulations of speeech production," unpublished Master's thesis, EE Department, UCLA, Los Angeles, CA, 1995.