Imaging applications in speech production research

Shrikanth Narayanan

AT&T Bell Laboratories, Murray Hill, NJ 07974

Abeer Alwan

Speech Processing and Auditory Perception Laboratory

Dept. of Electrical Engineering, UCLA, Los Angeles, CA 90095

ABSTRACT

The primary focus of speech production research is directed towards obtaining improved understanding and quantitative characterization of the articulatory dynamics, acoustics, and cognition of both normal and pathological human speech. Such efforts are, however, frequently challenged by the lack of appropriate physical and physiological data. A great deal of attention is, hence, given to the development of novel measurement/instrumentation techniques which are desirably non invasive, safe, and do not interfere with normal speech production. Several imaging techniques have been successfully employed for studying speech production. In the first part of this paper, an overview of the various imaging techniques used in speech research such as x-rays, ultrasound, structural and functional magnetic resonance imaging, glossometry, palatography, video fibroscopy and imaging is presented. In the second part of the paper, we describe the results of our efforts to understand and model speech production mechanisms of vowels, fricatives, and lateral and rhotic consonants based on MRI data.

Keywords: speech production, articulatory data, acoustic models, vocal tract geometry, area functions, tongue shapes, MRI.

1 INTRODUCTION

A significant part of research in human speech production focuses on understanding and modeling the dynamics of the various articulators such as the tongue, lips and the jaw during speech production, and on deriving relations between the vocal-tract configurations and the corresponding acoustic speech signals. Results of such physically- and physiologically-motivated modeling approaches have practical significance in applications such as the development of high-quality speech synthesizers, low-bit rate coders, and automatic speech recognizers.

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.

0-8194-2084-0/96/$6.00

Imaging techniques, both standard in clinical use and ingenious to speech research, have tremendously contributed to providing crucial details of the speech production apparatus.

In the first part of this paper, a brief overview of the various imaging techniques used in speech research such as x-ray radiography, ultrasound imaging, structural and functional magnetic resonance imaging, direct and indirect palatography, and video imaging is presented. In the second part of the paper, results of our efforts to understand and model speech production mechanisms of various speech sounds are described.

2 IMAGING THE HUMAN VOCAL TRACT

Several imaging techniques have been used in studying the details of the human speech production system. The term 'imaging' in this paper is used to refer to any technique that produces some form of visual/graphical information representation of the components or the functions of the human speech production system. For example, some of these techniques such as ultrasound and magnetic resonance can be used to observe the speech production apparatus more directly than some others such as palatography where one infers information about articulatory shapes indirectly from graphical images of tongue-palate contact patterns. Nevertheless, since most of these techniques essentially provide only partial or limited information about certain specific attributes of speech production, data are frequently collected using several of these techniques in parallel, if not simultaneously. Novel signal processing techniques are often required to process, visualize, and quantify the various physical and physiological speech production data thus obtained. Typically, the use of a particular technology (or a combination of several of those, if possible) is dictated by two major factors: (1) the particular part or region of the articulatory system under investigation. For example, lip shapes can be easily imaged directly using simple video techniques while the analysis of tongue shapes requires sophisticated ultrasound and/or magnetic resonance imaging. (2) the way the data are used: qualitative and/or quantitative analysis. For example, ultrasound images are more useful for qualitative analyses of tongue shapes while MRI data lend themselves to length, area, and volume measurements of the human vocal tract.

The other important factors in the popularity of a particular technology include its safety and ease of use. For example, one of the main reasons that MRI is more attractive than x-rays is that it does not involve any known radiation risks. However, due to limitations in the imaging speed, MRI which is non invasive, is currently not very useful to observe the larynx when compared to video fiberscopes which require insertion of an optical cable through the nasal passage into the pharyngeal region.

2.1 X-ray techniques

In the past, cine x-ray techniques had been popular in speech research to obtain lateral (midsagittal) images of the vocal tract.[1–5] Cross-sectional vocal tract areas (area functions) were estimated from these midsagittal images for acoustic modeling. The use of such radiographic techniques has been reduced significantly due to radiation risks and to the limited information obtained. The x-ray microbeam technique[6–8] reduces the radiation dosage by tracking just a few lead pellets located along the midsagittal plane but does not provide information about the vocal tract cross-sections or information about the pharyngeal region. Techniques that use magnetic field for tracking pellets mounted on articulators, such as the Electromagnetic midsagittal articulometer (EMMA),[9] have provided successful alternatives to radiation-based methods. Computer-aided tomography[10] is capable of yielding cross-sectional information but still suffers from radiation risks and relatively low speed of imaging. Ultrasound and MRI have proved to be viable alternatives for such investigations.

2.2 Ultrasound

Ultrasound technology has provided an acceptable mean for studying tongue shapes and movements during speech production.[11] The technology is safe and non invasive and the imaging speed is suitable for studying the dynamics of speech production. However, the entire vocal tract can not be studied by this method. Moreover, due to the presence of the airway above the tongue, the palate can not be imaged, while due to the presence of air space below, the tongue tip/blade can not be successfully captured. Nevertheless, modified ultrasound techniques have helped to further the understanding of the 3D model of tongue, either by using multiple scanning procedures[12] or by using other parallel instrumentation measurements such as x-ray microbeam[13] or palatography.[14]

2.3 MRI

Magnetic resonance imaging (MRI) is a powerful tool in obtaining the vocal-tract geometry and does not involve any known radiation risks. The images have good signal to noise ratio, are amenable to computerized 3-D modeling, and provide excellent structural differentiation. In addition, the tract (airway) area and volume can be directly calculated. The low image sampling rate, however, has restricted MRI use to the study of sustained speech sounds, corresponding to 'static' tract shapes. In addition, the high expense associated with using MRI equipment, has restricted its use in speech research. Previous MRI studies have been limited to static configurations of vowels[15–17] and consonants.[18–20] Recent advances in Echo-planar imaging are aimed at alleviating the problem of the low-image sampling rate and allow imaging of dynamical sequences. To our knowledge, no study has attempted to explore faster MRI imaging techniques, such as echo-planar imaging, to image the dynamical vocal-tract configurations during speech production. Recently, there has also been considerable interest in understanding higher level processing during speech production by monitoring brain activity using functional MRI by itself or in conjunction with techniques such as positron emission tomography (PET) and magnetic encephalography (MEG).

2.4 Palatography

In palatography, the contact of the tongue with the palate, alveolar ridge, and inner margins of the teeth is graphically registered. In static palatography,[21] a deposit material such as carbon powder is coated on the tongue surface prior to speaking and the resulting contact patterns are captured through video imaging. The limitation of this method is that it yields only the maximum contact pattern. The dynamical linguopalatal contact during most consonant and certain vowel productions provides insights into the articulatory dynamics and can be obtained through dynamic electropalatography (EPG). The way EPG works is that the tongue contact on the palate is registered on an array of contact sensing electrodes mounted on a (custom-fitted) pseudo-palate in the subjects' mouth.[22,23] The raw EPG (palatal contact) data can then be processed for studying the various specific features such as total contact, symmetry of contacts, constriction width and location. A sequence of EPG data frames over the course of the consonant production reveals the temporal variation of the tongue contact patterns.

2.5 Other methods

Video imaging is sometimes used in speech-production research. Dynamics of externally-visible articulators such as the lips can be captured by video imaging.[24] Knowledge of the dynamics of such external articulators is also important for understanding audiovisual interaction in multimodal organization of speech perception. Internal regions of the vocal tract such as the larynx can be imaged by using a fiberscope[25] wherein an optical fiber (or, two) is inserted through the nasal passage into the pharynx. The laryngeal actions can then be imaged through conventional video techniques. The obvious disadvantage is the invasive nature of this technique. Glossometry

is a technique which uses LED-photosensor pairs mounted along the midsagittal line of a pseudopalate to assess tongue postures and shapes.[26] This technique has also found use in speech training of the hearing-impaired by means of a visual feedback of the articulatory functions.

In the next section of the paper, recent results of the application of MRI and EPG techniques in the articulatory-acoustic analysis and modeling of vowel and consonant sounds is presented.

3 ARTICULATORY-ACOUSTIC ANALYSIS AND MODELING: CASE STUDY

The data we collected and analyzed are Magnetic Resonance Images (MRI), acoustic recordings, and Electropalatography (EPG) data from four phonetically-trained native talkers of American English during the sustained production of certain sounds. The MR images were useful for characterizing the 3D geometry of the human vocal tract. The data were also used for providing measurements of lengths, area functions, and volumes of the vocal tract and other cavity structures such as the piriform sinuses and the sublingual cavities. EPG was used to study inter- and intra-speaker variabilities in the articulatory dynamics. High quality recordings and other aerodynamic experiments were used to provide the acoustic and aerodynamic data needed for the modeling. We will illustrate inter and intra-speaker characteristics of vocal-tract and tongue shapes for various speech sounds and show our results of acoustic modeling based on the MRI and acoustic data for a subset of these sounds; namely, fricative consonants.

3.1 Subjects

Four phonetically-trained, native American English speakers [2 males (MI, SC) and 2 females (AK, PK)] served as subjects. Subjects AK and MI, both in their twenties at the time of the experiments, were raised in Northern California and have spent the seven years preceding this study in Southern California. Subject SC, in his thirties, spent the first ten years of his life in Indiana and has since been in California. Subject PK, in her early forties, lived in New Jersey and Ohio her first three years, and in the Boston area through her thirties. Since then she has been in the Los Angeles area.

3.2 Magnetic resonance imaging (MRI)

A detailed description of the acquisition and analysis procedures is provided in Narayanan *et al.*[19] Magnetic resonance (MR) images were collected using a GE 1.5 Tesla SIGNA machine with a fast SPGR (radio frequency spoiled GRASS) protocol in the coronal, axial, and sagittal planes. The image slice thickness was 3 mm with no interscan spacing. Each image was represented by a 256 X 256 pixel matrix, yielding a resolution of 0.0081 cm^2 per pixel for an FOV = 24 cm. The subjects, in supine position, sustained each consonant for about 13-16 s enabling four to five image slices to be recorded. The consonants were produced in a neutral vowel context. A special head-neck coil, (by *Medical Advances*), which helped maintain the subjects' heads in a fixed position, was used to enhance the SNR of the images.

The scanning region for the coronal and axial planes included the region between the lips and the posterior pharyngeal wall along the antero-posterior axis and the region between the top of the hard palate and just below the eighth vertebra along the infero-superior axis. Coronal and axials scans were taken approximately perpendicular to the vocal-tract midlines, in the buccal and pharyngeal regions, respectively, based on a midsagittal

localizer image for each subject. Similarly, the scanning region for the sagittal plane was based on axial and/or coronal localizer images. The data set comprised 28 to 35 images/sound/subject in the sagittal plane, and 40 to 45 images/sound/subject in the axial and coronal planes. In addition, reformatting of the raw images was used to obtain cross-sections along any desired (oblique) plane. For example, area information along the vocal tract's bend was obtained by image reformatting. Since midsagittal profiles provide the most convenient reference for specifying grid locations for performing area calculations, sagittal scans are chosen for area calculations along the vocal tract bend from reformatted images. Midsagittal data are also used for length measurements.

Automatic segmentation of the vocal-tract regions in the images was followed by a careful manual verification of the selected regions in each image. Following segmentation, three-dimensional reconstructions of the entire vocal tract, or specific regions such as sublingual cavities, could be made by computer-aided concatenation of the selected regions of interest. Length, area, and volume measurements could be made directly using a pixel counting algorithm.

Articulatory analysis and measurements were performed in several steps. First, overall vocal-tract and tongue shapes were analyzed using raw images and complete 3D models reconstructed from appropriately segmented raw scans. All the 3D reconstructions reported in this study were constructed using coronal scans. Analysis of the buccal region is primarily based on sagittal and coronal sections while that of the pharyngeal region is based on sagittal and axial profiles. In addition, interactive slicing of the 3D objects (along any desired plane) using image processing software facilitated the morphological analyses.

Area measurements were made in two stages: in the first stage, cross-sectional areas were directly measured from the coronal and axial scans to provide information on the front (buccal) and back (pharyngeal/laryngeal) regions, respectively; in the second stage, sagittal scans were reformatted to obtain areas along the planes perpendicular to the midline of the vocal tract along the bend. To enable comparative graphical analyses across the various sounds and subjects, a simplified representation of the area function is considered. Areas up to the laryngeal inlet, defined by the section showing the complete separation of the piriform sinuses by the inter-arytenoid eminence, were considered. Furthermore, the "effective" area of the airway was obtained by a simplification of the morphology: subtracting tissues areas, such as the uvula, and the various epiglottal folds, from the total pharyngeal cavity areas.

3.3 Electropalatography (EPG)

EPG data from the subjects were recorded on a later date using *Kay Elemetrics Palatometer.* Each subject has a custom-fitted acrylic palate with 96 sensing electrodes. The sweep rate of this system is 1.7ms and the sampling period is 10 msec. The subjects assumed a supine position (similar to that assumed inside the MRI machine) while recording. The data for each subject were collected in a single session, that lasted about one and a half to two hours. The data were collected over a month (post-MRI experiments). The speech material consisted of vowels /a, i, u/, fricatives, lateral approximants (dark and light allophones) and rhotic approximants (in word-initial and syllabic positions), sustained for about 2-2.5 sec/token. Eight repetitions of each condition were obtained. The sounds were produced at a normal 'conversational' rate and level and, the consonants were preceded by the neutral vowel /ə/. For the purposes of this study, the total electrode region covered by the electrodes was broadly divided into several regions as illustrated in Figure 1. The percentage of electrodes that are contacted in each of these regions served as a basis for our analysis. Inter- and intra-subject variabilities in the linguopalatal contact profiles were studied using repeated-measures multifactorial ANOVA. It has to be noted, however, that EPG measures just the linguopalatal contact and is limited to the region between the teeth and the anterior part of the velum.

3.4 Vowels

MRI, EPG, and acoustic data were collected for the three-point vowels /a, i, u/ from the four subjects. Although our study focussed on just /a, i, u/, it offers many advantages when compared to other recent studies on vowels.[15–17] Our study included both male and female subjects, and had extensive MRI data (in all 3 anatomical planes). EPG data were also included in the study to supplement the MRI data to enable variability analyses. 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 recording of speech). In addition to providing 3D vocal tract renditions and length/area/volume measurements, our analysis includes 3D tongue shape analysis. The articulatory data are used in acoustic modeling, in both 1D and 2D simulations of acoustic wave propagation in the vocal tract. Finally, the vowel data form a part of a larger articulatory-acoustic data set that includes consonants such as fricatives, laterals, and rhotics from the same subjects.

Midsagittal images for /a, i, u/ show the vocal tract with significantly greater clarity than when compared to lateral x-ray images (Figure 2, male subject). Moreover, the axial and coronal cross-sections have helped reveal the tongue shape and the cavities in the pharyngeal region in much greater detail. 3D vocal tract and tongue shapes for /a, i, u/ (Figure 3, male subject) help to clearly demonstrate the differences between the three vowels. The anterior tongue body is convex and closer to the palate in /i/ while it is flat or slightly concave and farther away from the palate in /a/. On the other hand, /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 relatively smaller than in /i/. These vocal tract characteristics explain the area functions shown in Figure 4.

3.5 Fricatives

The data included all eight fricatives in English: the stridents: /ʃ, s, ʒ, z/, and the nonstridents:/θ, f, ð, v/. A detailed morphological description of the vocal tract using MRI data is given in Narayanan *et al.*[19] A sample 3D vocal tract for the fricative /s/ of a male subject is shown Figure 5. Sample 3D tongue shapes and linguopalatal contact profiles for the fricatives /s/ and /ʃ/ of the same male subject are shown in Figure 6 and Figure 7, respectively. The vocal tract cross-sections were found to be best approximated by elliptical/semi-elliptical shapes. Results showed similar vocal tract shapes across subjects for each sound (place of articulation). In general, among the fricatives, the labiodentals exhibited the most variability across speakers. The voiced fricatives showed tendency towards tongue root advancement and hence, resulting in relatively larger pharyngeal volumes. The anterior tongue body was concave shaped with medial grooving in /s, z/ while it was convex and relatively high positioned in /ʃ, ʒ/. The degree of grooving was speaker dependent with a definite correlation between the degree of grooving and lateral linguopalatal contact (as seen through EPG). Asymmetries in tongue shapes and linguopalatal contact patterns were subject, and perhaps sound, dependent and not on the supine position assume while speaking.

Area functions measured from the MRI data were used for acoustic modeling of these sounds.[27,28] For illustration, area functions for the strident fricatives /s, z/ and /ʃ, ʒ/ of a male subject are shown in Figure 8. As a first approximation, the vocal tract is modeled as a concatenation of uniform cylindrical tube-sections each section being 3 mm long. Depending on the subject's vocal-tract length, the total number of sections is 55-60. Sublingual cavities, such as those present in /ʃ/ and /ʒ/ are modeled as shunt branches specified in the anterior buccal cavity. Once the area function is known, several approaches may be used for simulating the acoustics in the vocal tract. In the current study, a time-domain simulation method proposed by Maeda[29] was used to determine the vocal-tract transfer function for each sound. Source models for fricative consonants were then derived based on aerodynamic principles of sound generation, in conjunction with vocal-tract models obtained from MRI data. Results indicate that a linear source-filter model is adequate for capturing essential spectral characteristics of sustained fricatives below 10 kHz. The results comparing natural and synthesized spectra for

the fricatives /s/ and /ʃ/ (corresponding to the area functions of Figure 8) are shown in Figure 9. The hybrid source models used employ a combination of acoustic monopole and dipole sources, and a voiced source in the case of the voiced fricatives. The number of sources, source locations and spectral characteristics are chosen based on an analysis-by-synthesis approach and are motivated by aeroacoustic theory. The resulting model is computationally efficient and can be readily used for synthesis.

3.6 Lateral approximants

Articulatory and acoustic data corresponding to the dark and light allophones of the lateral approximant /l/ in American English (denoted by [l] and [ɫ], respectively) were collected. MR images for both [l] and [ɫ] indicate that the *midsagittal* tongue contours can be different across subjects. Common characteristics, however, were revealed in cross-sectional and 3D tongue shapes, area functions, and linguopalatal contact profiles. These sounds were characterized by a complete linguo-alveolar contact or, just a constriction as observed in the [ɫ] of one subject. The '*lateral channels*' along the sides of the tongue began appearing from where the alveolar occlusion/constriction was seen and continued posteriorly until lingua-velar contact was established (5-6 cm from the lip opening). The right and left channels appear to be, in general, unequal and their areas start increasing behind the alveolar occlusion (due to inward lateral compression of the tongue body) and start decreasing again as the region of lingua-velar contact is approached. Analysis of the overall 3D tongue shapes revealed that the overall tongue body shape behind the occlusion tends to be convex. The 3D tongue shapes indicate that the posterior tongue body shows a tendency towards an *inward lateral compression* which is directed towards the midsagittal plane. This enables the creation of lateral flow channels in the space between the curved sides of the tongue body and the teeth. In addition, a 'grooving' tendency along the midsagittal line was observed in some subjects, particularly in the region behind the alveolar contact. This anterior medial grooving observed in the laterals, which is less prominent than that observed in alveolar sibilants such as /s/,[19] is attributed to the secondary effect of inward compression of the posterior tongue body. Unlike alveolar fricatives, the grooving does not continue through the posterior tongue region as a concave surface, suggesting that it is not a key component of a medial airflow channel. Asymmetry in tongue shapes and linguopalatal contact profiles were found to be subject-dependent. Results of statistical analyses using EPG data, however, indicated no significant asymmetry effects. Comparison of [l] and [ɫ] showed greater linguopalatal contact in [l] than in [ɫ]. The areas in the region behind the alveolar contact were smaller in [l] than in [ɫ]. Consistently small pharyngeal areas were found in [ɫ] due to tongue-root retraction and/or posterior tongue body raising.

3.7 Rhotic approximants

During the production of the American English /r/, the vocal tract appears to be characterized by three cavities due to the presence of two distinct supraglottal constrictions. The primary constriction occurs in the buccal cavity and the secondary constriction, in the pharyngeal cavity. A sample 3D vocal tract of /r/, by a male subject (SC), and the 3D tongue shape associated with it are shown in Figure 10. The buccal constriction may occur anywhere in the palatal region: the more forward locations are typically due to a raised anterior tongue and the posterior ones, due to a raised dorsum. For our subjects, the buccal constriction began anywhere between 2.4-5.4 cm away from the lips and extended over 1.5-2.5 cm with minimum areas ranging between 0.25-0.7 cm^2. The secondary constriction occurs typically in the mid-pharyngeal region due to an advanced tongue root ('pharyngealization'). Analyses indicated that a more anterior buccal constriction was associated with a more superior pharyngeal constriction. A large volume anterior to the buccal constriction resulted due a tongue body that was drawn inwards, away from the lips. The anterior tongue body was characterized by convex cross-sections. Similarly, a large volume posterior to the buccal constriction (and superior to the pharyngeal constriction) was created by a significantly lowered posterior tongue body that exhibits a prominent concave shaping. The change in the cross sections, from the convex anterior shapes to the more concave posterior shapes, appeared to be more abrupt for the buccal constrictions that were at a more posterior location, resulting in more abrupt changes in

the area functions. These observations suggest an interplay between the relative locations of the buccal and pharyngeal constrictions and the 3D tongue shapes. Variabilities in the details of the relative cavity sizes and their locations, which largely depend on the individual subject's articulation patterns and oral morphology, are expected to introduce variabilities in the corresponding acoustic patterns.

Subjects AK and MI produced /**r**/s as they would appear in 'word-initial' and 'syllabic' positions while PK deliberately produced tongue tip-up and bunched /**r**/s. Subject SC produced only the 'word-initial' version. Comparison of the bunched and tip-up /**r**/s produced by PK revealed that, in spite of the raised tongue tip in the latter case, the primary buccal constriction is attributed to a raised dorsum in both cases, and a three-cavity vocal tract description still holds. For the other subjects, the general tongue body shapes and area functions appeared very similar for the /**r**/s in both word-initial and syllabic positions although syllabic /**r**/s tended to show larger areas in the cavity between the buccal and pharyngeal constrictions. The buccal constriction for AK's and MI's /**r**/s were produced with a raised dorsum resulting in a tongue body shape that resembles a canonical bunched /**r**/. The /**r**/ of SC, on the other hand, was produced with a raised anterior tongue body, rather than a raised tongue tip, resulting in a more anterior, and shorter, buccal constriction when compared to those seen in the other subjects. The results of this investigation may be used as a baseline for studying articulatory-acoustic relations of these sounds.

4 ACKNOWLEDGMENTS

The authors would like to thank Dr. Katherine Haker and members of the Cedars Senai Imaging Medical Center, their subjects and other members of the UCLA Phonetics Laboratory for their assistance and support. This work was done while Dr. Narayanan was a Ph.D. student at the UCLA Speech Processing Laboratory. Research supported by NSF.

5 REFERENCES

[1] G. Fant, *Acoustic Theory of Speech Production.* Mouton: The Hague, 1960.

[2] J. S. Perkell, *Physiology of Speech Production: Results and Implications of a quantitative cineradiographic study.* Cambridge, MA: MIT, 1969.

[3] J. D. Subtelny, N. Oya, and J. D. Subtelny, "Cineradiographic study of sibilants," *Folia Phoniatrica*, vol. 24, pp. 30–50, 1972.

[4] S. B. Giles and K. L. Moll, "Cineflurographic study of selected allophones of English /l/," *Phonetica*, vol. 31, pp. 206–227, 1975.

[5] P. Delattre and D. C. Freeman, "A dialect study of American r's by x-ray motion picture," *Linguistics, An international review*, vol. 44, pp. 29–68, 1968.

[6] O. Fujimura, "Medical implications of on-line computer experiments in speech research," *Tokyo J. Med. Sci.*, vol. 75, pp. 235–239, 1967.

[7] O. Fujimura, S. Kiritani, and H. Ishida, "Computer-controlled radiography for observation of articulatory and other human organs," *Comp. Biol. Med.*, vol. 3, pp. 371–384, 1973.

[8] S. Kiritani, K. Itoh, and O. Fujimura, "Tongue-pellet tracking by a computer-controlled x-ray microbeam system," *J. Acoust. Soc. Am.*, vol. 57, no. 6, pp. 1516–1520, 1975.

[9] J. S. Perkell, "Electromagnetic midsagittal articulometer systems for transducing speech articulatory movements," *J. Acoust. Soc. Am.*, vol. 92, no. 6, pp. 3078–3096, 1992.

[10] S. Kiritani, Y. Tateno, and T. Iinuma, "Computer tomography of the vocal tract," in *Dynamic aspects of Speech Production* (M. Sawashima and F. S. Cooper, eds.), pp. 203–206, University of Tokyo press, 1977.

[11] E. Keller and D. Ostry, "Computerized measurement of tongue dorsum movement with pulsed echo ultra sound," *J. Acoust. Soc. Am.*, vol. 73, pp. 1309–1315, 1983.

[12] K. L. Watkin and J. M. Rubin, "Pseudo-three-dimensional reconstruction of ultrasonic images of the tongue," *J. Acoust. Soc. Am.*, vol. 85, pp. 496–499, 1989.

[13] M. Stone, "A three-dimensional model of tongue movement based on ultrasound and x-ray microbeam data," *J. Acoust. Soc. Am.*, vol. 87, pp. 2207–2217, 1990.

[14] M. Stone, A. Faber, L. J. Raphael, and T. H. Shawker, "Cross-sectional tongue shapes and linguopalatal contact patterns in [s], [S] and [l]," *J. Phonetics*, vol. 20, no. 2, pp. 253–270, 1992.

[15] 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.

[16] 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.

[17] A. R. Greenwood, C. C. Goodyear, and P. A. Martin, "Measurement of vocal tract shapes using magnetic resonance imaging," *IEE Proc. I (Commun., Speech Vision)(UK)*, vol. 139, no. 6, pp. 553–560, 1992.

[18] J. Dang, K. Honda, and H. Suzuki, "MRI measurements and acoustic investigation of the nasal and paranasal cavities," *J. Acoust. Soc. Am.*, vol. 94, no. 3, p. 1765 (A), 1993.

[19] S. S. Narayanan, A. 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.

[20] S. Narayanan, A. Alwan, and K. Haker, "An articulatory study of liquid approximants in american english," in *Proc. of the XIII Intnl Congress of Phonetic Sciences*, (Stockholm, Sweden), 1995.

[21] P. Ladefoged, "Use of palatography," *J. Speech Hear. Dis.*, vol. 22, pp. 764–774, 1957.

[22] S. Fletcher, M. McCutcheon, and M. Wolf, "Dynamic palatometry," *J. Speech and Hearing Research*, vol. 18, pp. 812–819, 1975.

[23] W. J. Hardcastle, W. Jones, C. Knight, A. Trudgeon, and G. Calder, "New developments in electropalatography: A state of the art report," *Clinical Linguistics and Phonetics*, vol. 3, pp. 1–38, 1989.

[24] K. Honda, T. Kurita, Y. Kakita, and S. Maeda, "Physiology of the lips and modeling of the lip gestures," *J. Phonetics*, vol. 23, pp. 243–254, 1995.

[25] M. Sawashima, "Fiber optic observation of speech organs," in *Dynamic aspects of Speech Production* (M. Sawashima and F. S. Cooper, eds.), pp. 31–46, University of Tokyo press, 1977.

[26] S. G. Fletcher, M. J. McCutcheon, S. C. Smith, and W. H. Smith, "Glossometric measurements in vowel production and modification," *Clinical Linguistics and Phonetics*, vol. 3, pp. 359–375, 1989.

[27] S. S. Narayanan, *Fricative consonants: An articulatory, acoustic, and systems study.* PhD thesis, UCLA, Dept. of Electrical Engineering, Los Angeles, CA, June 1995.

[28] S. Narayanan and A. Alwan, "Parametric hybrid source models for voiced and voiceless fricative consonants," in *IEEE Proc. ICASSP*, (Atlanta, GA), p. To appear, May 1996.

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

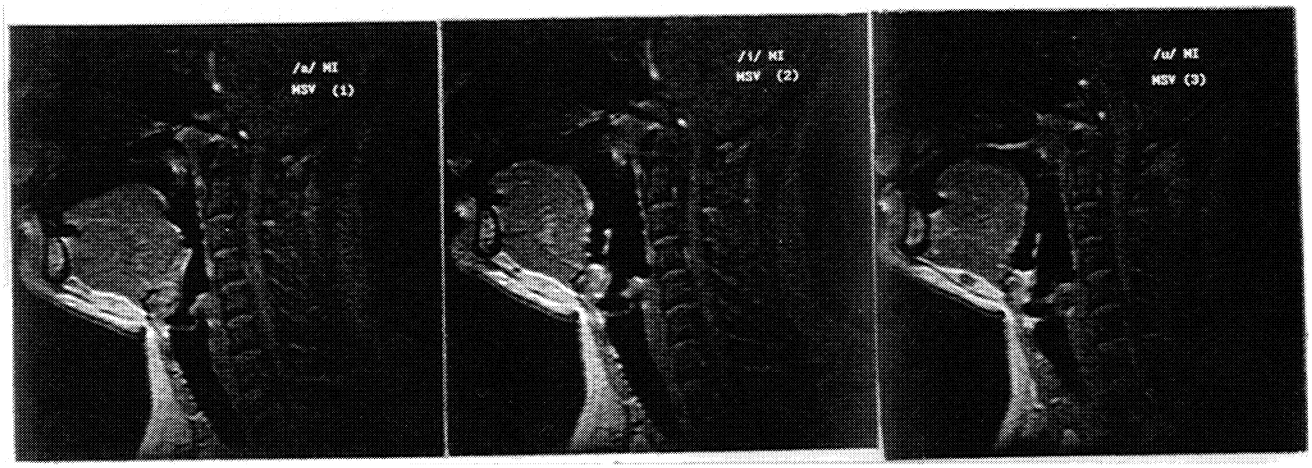


Figure 2: *Midsagittal vocaltract images for the vowels /a,i,u/ of a male subject.*

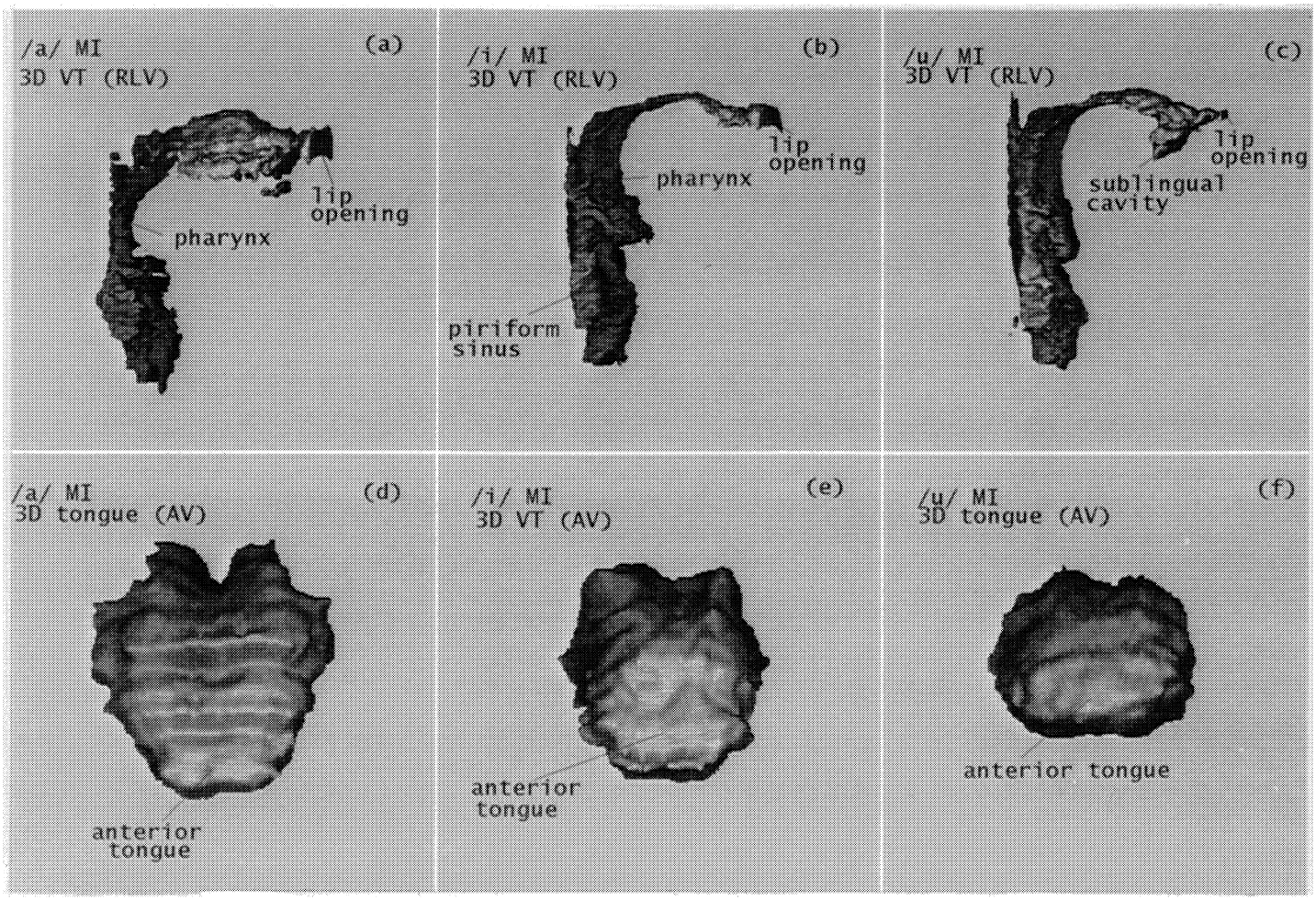


Figure 3: *3D vocal tract and tongue shapes for the vowels /a,i,u/ of a male subject (MI)*

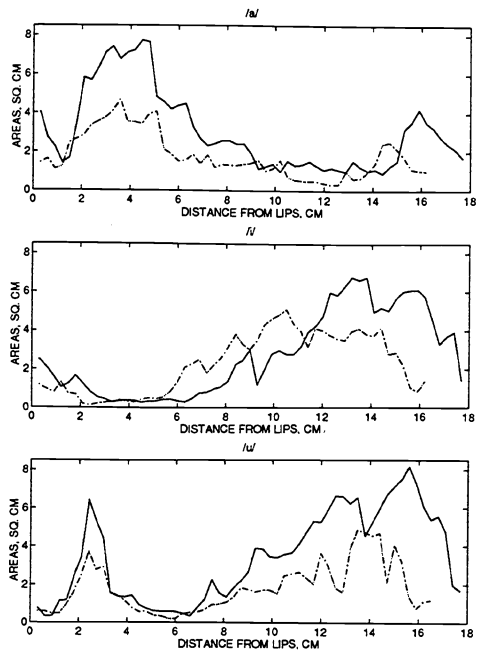


Figure 4: *Area functions for vowels: solid (male), dot-dashed (female).*

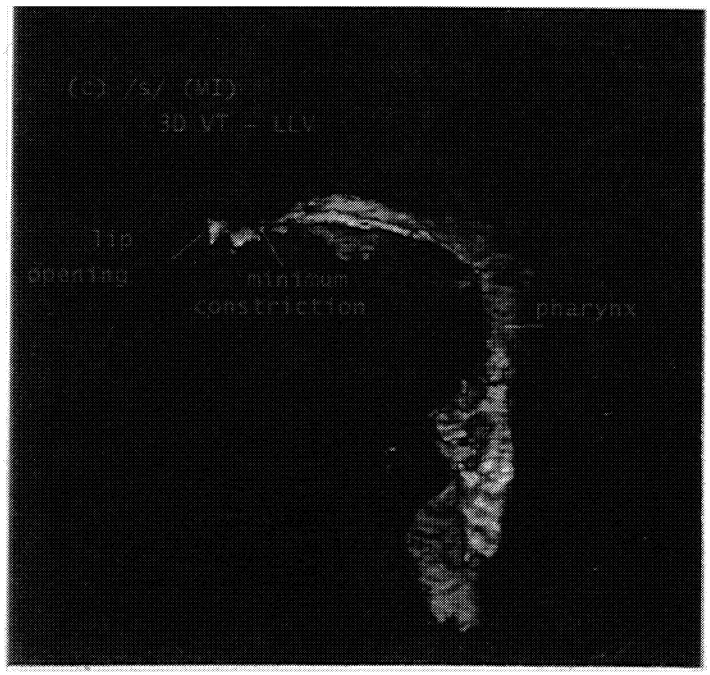


Figure 5: *Lateral view of the 3D vocal tract for the fricative /s/ for a male subject.*

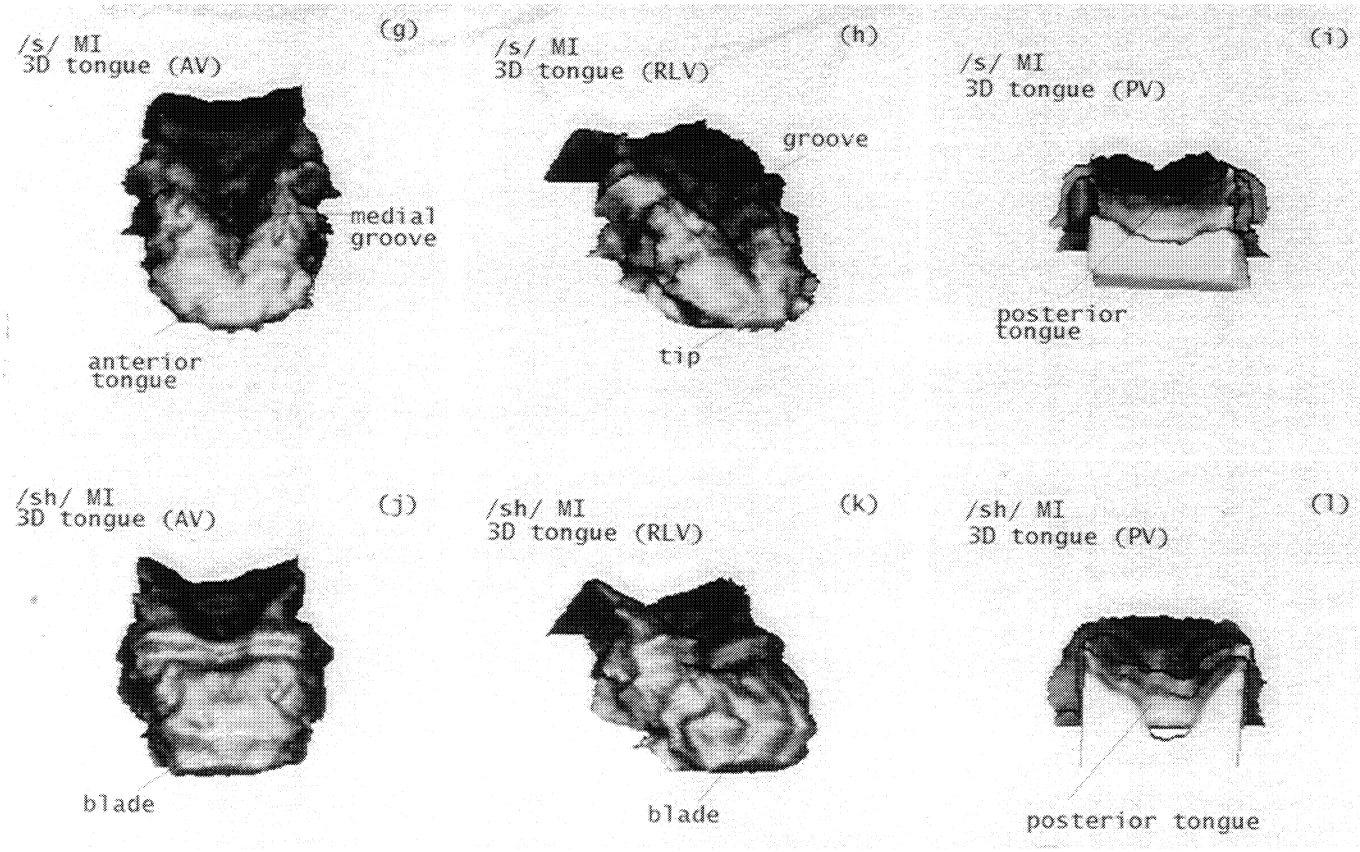


Figure 6: *3D tongue shapes for the fricatives /s/ and /sh/ for a male subject.*

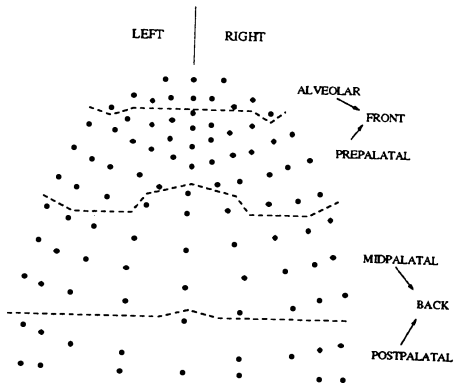
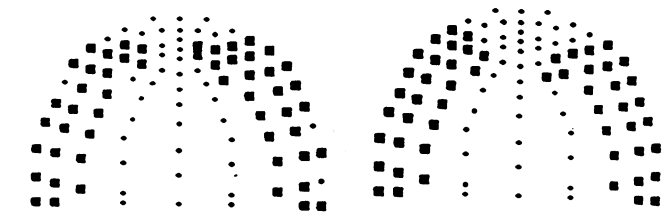


Figure 1: *Schematic of electrode array on the pseudo-palate.*

Figure 7: *Sample linguopalatal contact profiles for /s/ and /ʃ/.*



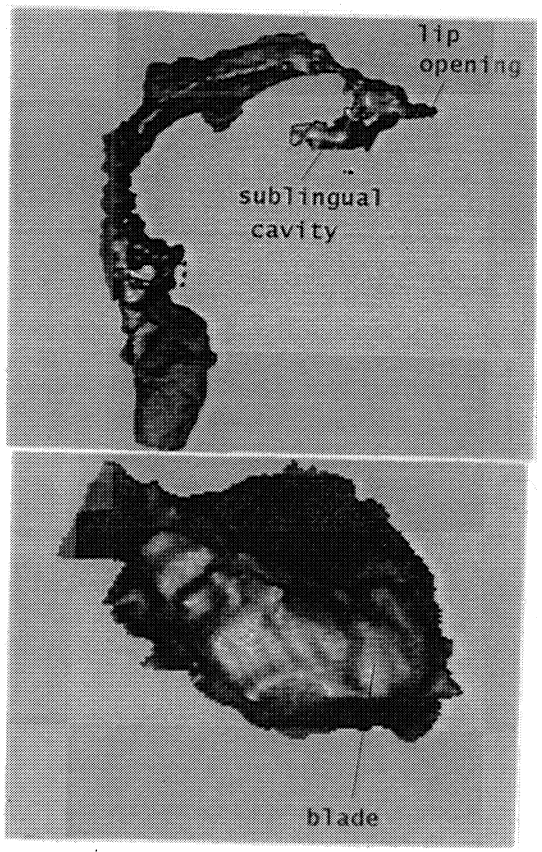


Figure 10: *3D vocal tract and tongue shape for /r/.*

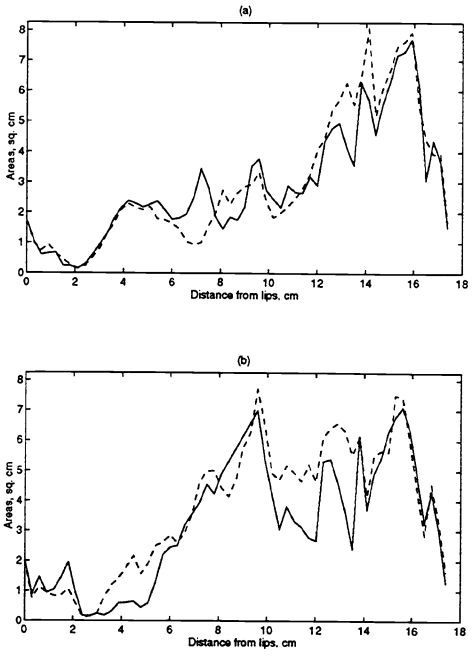


Figure 8: *Area functions for strident fricatives of a male subject MI (a) /s, z/, (b) /ʃ, ʒ/: unvoiced (solid), voiced (dashed).*

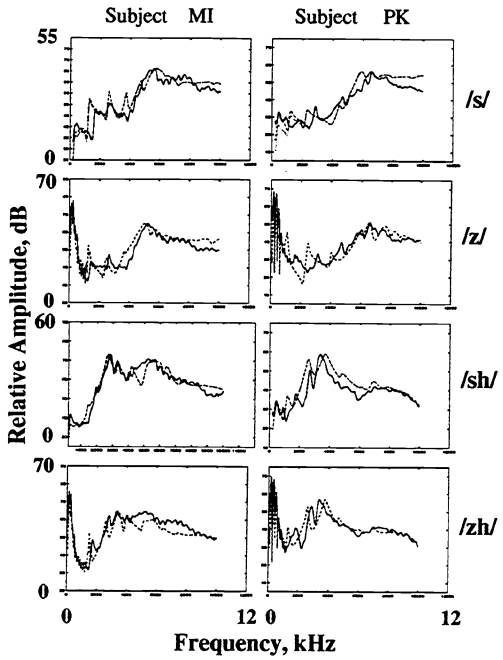


Figure 9: *Modeling results for strident fricatives: synthesized spectra (dashed), natural spectra (solid).*