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

AN MRI STUDY OF FRICATIVE CONSONANTS

Shrikanth Narayanan[1], *Abeer Alwan*[1] *and Katherine Haker*[2]
UCLA – 405 Hilgard Avenue, Los Angeles, CA 90024
[1]Electrical Engineering Department
[2]Dept. of Radiology and Cedars-Sinai Medical Center

10.21437/ICSLP.1994-159

ABSTRACT

Magnetic resonance images of the vocal tract during sustained production of the fricatives /s, ʃ, f, θ, z, ʒ, v, ð/ by four subjects are analyzed. Measurements of vocal tract lengths and area functions and analysis of the tongue shapes for different sounds are presented. Inter-speaker differences in area functions are found to be greater in the pharyngeal cavity than in the buccal cavity with the nonstrident fricatives exhibiting greater differences than the strident ones. The concave tongue shapes of /s/ and /z/ result in a more abrupt area function behind the constriction when compared to that of /ʃ/ and /ʒ/. Voiced fricatives show larger pharyngeal volumes than the unvoiced fricatives due to tongue-root retraction.

1. INTRODUCTION

Knowledge of the 3D geometry of the human vocal tract (VT) is an important factor in modeling the production mechanisms of speech sounds. Magnetic resonance imaging (MRI) is a powerful tool in obtaining this geometry and does not involve any known radiation risks. The images have good signal to noise ratio (SNR) and are amenable to computerized 3-D modeling of the VT. The low image sampling rate, however, has restricted the use of MRI to the study of sustained speech sounds. Most previous MRI studies have been limited to vowels (for example, [1]). In this paper, a morphological analysis of the VT geometry obtained from MR imaging in axial, coronal and sagittal planes of the fricatives /s, ʃ, f, θ, z, ʒ, v, ð/ is reported. Issues such as inter-speaker variabilities and voiced-unvoiced differences are addressed.

2. METHOD

MR images were collected using a GE 1.5 Tesla SIGNA machine with a fast SPGR protocol (TE = 4.0 msec, TR = 12.6 msec, NEX=2, FOV=20 or 24 cm). The image slice thickness was 3 mm with no interscan spacing. The subjects assumed a supine position inside the scanner. A special head-neck coil, which helped keep 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, chosen based on a mid-sagittal localizer image for each subject, 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. 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 for the axial and coronal planes. In this study, coronal and axial scans are used to measure and analyze the cross-sectional areas and morphology of the "front" region (buccal cavity extending from the lips to the posterior pharyngeal wall) and "back" region (pharyngeal cavity extending from the uvula to the beginning of the trachea), respectively. Sagittal scans are used for area function measurements, particularly along the VT bend. Areas obtained from sagittal scans are contrasted with those obtained from coronal and axial scans. Mid-sagittal data are also used for length measurements.

Four phonetically-trained, native American English speakers [2 males (MI, SC) and 2 females (AK, PK)] served as subjects. During scanning, the speakers sustained each fricative for about 13-16 sec enabling four to five image slices to be recorded (about 3.2 sec/image). The consonants were anticipated in a VC context with the neutral vowel /ə/. The subjects repeated each sound six to nine times, with a pause of three to ten seconds between repetitions, to enable the entire VT to be scanned.

3. ANALYSIS

The data were processed on an ISG-Allegro Silicon Graphics workstation equipped with software for MR image processing. The first step in data processing involved segmenting the regions of interest in the image. Preliminary segmentation of the VT airway was achieved using an automatic thresholding procedure that relies on the contrast between the airway and the surrounding tissues. Due to variabilities and complexities in the vocal tract morphology (for example, the lower-pharyngeal and laryngeal regions) and non-uniformities in the image contrast (for example, regions near the teeth) automatic thresholding procedures may result in erroneous segmentation. Hence, automatic segmentation was followed by a careful verification of the selected regions in each image and appropriate boundary corrections were made manually. Various aids such as radiology/anatomy atlases, dental casts of the subjects were used to ensure accurate segmentation. Following segmentation, three dimensional reconstruction of the entire VT, or specific regions such as sublingual areas, could be made by computer-aided concatenation of the selected regions of interest. Length, area, and volume measurements could be made directly from any specified region of interest using a pixel counting algorithm.

3.1 Area function and length measurements:

Although coronal and axial scans can provide accurate area estimates of the front and back regions, respectively, area functions (AF's) along the VT bend can not be directly measured from these scans. Hence, in this study, area measurements are performed at two levels: in the first level, "raw" cross-sectional areas are directly measured from the coronal and axial scans to provide information on the front and back regions, respectively; in the second level, sagittal scans are reformatted to obtain AF's along the planes perpendicular to the midline of the VT including the VT bend. Our calibration experiments indicate that, in general, over-estimated errors in area measurements range between 2-15%; for volume measurements they are 1.5-8%, and for lengths, errors range between 1.5-8.5%. In general, measurement of smaller dimensions results in larger errors. Similarly, measurements made from reformatted sections involve larger errors when compared to those made directly from raw scans.

3.2 Coronal measurements: Dimensional measurements in the front region yield important information regarding the constriction geometry and cross-sectional tongue shapes. Sections in the buccal cavity beginning at the lips and ending near the posterior edge of the back of the tongue body, showing no pharyngeal connections, are considered for area function measurements in the front region. Estimates of the area of minimum supraglottal constriction (A_C) are made directly from raw coronal scans. Although the image slice corresponding to A_C gives an approximate indication (within ±3 mm) of the location of the constriction (l_C), a more accurate l_C estimate could be made from the mid-sagittal profile. Areas of the sub-lingual cavities, such as those evident in the productions of /ʃ/ and /ʒ/, and detailed analyses of the cross-sectional tongue shapes are also made with coronal scans. Two main problems are faced in the area calculations of the front region: (1) There are difficulties in specifying the boundaries at the regions sideways to the lips, typically in the first section through the lips. In most cases, however, approximate boundaries could be specified based on the upper and lower lip boundaries. (2) Unclear teeth boundaries with respect to the airway also poses problems, particularly in the anterior part of the front region. Segmentation at the teeth boundaries is carefully made by following the gingival outline and employing knowledge of each subject's oral structure.

3.3 Axial measurements: Areas in the back region are calculated from the first effective slice in the uvular region (which appears distinct from the buccal cavity) to the beginning of the trachea (typically at the mid-level of the inter-vertebral disk between the fifth and sixth vertebra). The area calculations in the back region pose problems due to the appearance of several cavities because of the presence of tissue structures such as the uvula, epiglottis, and vocal folds, along the VT air way. In an attempt to make the area calculations in a systematic manner, the back region is divided into several zones such as the *uvular, upper-pharyngeal, lower-pharyngeal, laryngeal* and *subglottal* regions. Areas of each cavity such as the epiglottic valleculae and piriform sinuses, are measured separately. In this paper only a simplified representation of the back region AF's is presented. Areas up to the laryngeal inlet, defined by the section showing the complete separation of the piriform sinuses by the inter-arytenoid eminence, are considered. Furthermore, the "effective" area of the airway is obtained by a simplification of the morphology: subtracting the areas of the tissue, such as the uvula, and the various epiglottal folds, from the total pharyngeal cavity areas. A detailed description of the pharyngeal region will be presented in a future publication.

3.4 Sagittal measurements: Sagittal scans are used for the following: (a) Mid-sagittal profiles are used for various length measurements such as the length of the vocal tract (l_{VT}), length of the front and back regions (l_{FR} & l_{BR}), upper and lower lip protrusions (l_{PU} & l_{PL}), lateral lip opening (l_{LO}), and location of the minimum supraglottal constriction (l_C). (b) Sagittal scans are reformatted to obtain cross-sections along planes orthogonal to the midline of the vocal tract and used for area function calculations. In addition, this option of interactively obtaining cross-sections at any specified location, and along any arbitrary plane, serves as a valuable tool in the analysis of the cross-sectional morphology. (c) Mid-sagittal MRI data are used for comparative analyses with published mid-sagittal X-ray data.

4. RESULTS AND DISCUSSION

4.1 Front region: Analysis of the front region is based on coronal scans and midsagittal profiles. Sample front region area functions for /s/, /z/, and /f/ are shown in Fig. 1(a)-(c). Analysis reveals consistency in the general VT shapes of the front region across speakers for each fricative with greater similarities in the case of the strident fricatives /s, ʃ, z, ʒ/. The maximum measured area in the front region is 7.26 cm^2 (/f/ of MI). Values of l_C and A_C are listed in Tables I and II, respectively. The measured l_C values are comparable to those derived from X-ray data [2].

The front region VT shapes are, in general, similar for the voiceless and voiced fricatives that share the same place of articulation. The voiced fricatives, however, tend to have slightly larger AF values in the region behind the constriction when compared to their voiceless counterparts. The posterior tongue region for the voiced fricatives is raised slightly higher than it is for the voiceless fricatives, suggesting the influence of the tongue root retraction evident in the voiced cases. The l_C values for the voiced and unvoiced sounds are more similar than the A_C values. Strident fricatives, in general, are characterized by smaller A_C values than the non-stridents. Variability in the A_C values implies variability in the aerodynamics, such as the degree of turbulence, and varying degrees of coupling between the back and front cavities.

Strident fricatives

Constriction region: The cross-sectional shape at the constriction for the stridents, in general, resembled a slit, rather than a circular orifice, due to a flat tongue surface at the constriction. Smaller constriction widths, together with marked central tongue grooving, however, may result in a more circular orifice-like constriction, as is revealed in the /s/ and /z/ of PK. The constriction for /ʃ/ and /ʒ/ occurs in the postero-alveolar/antero-palatal region, with the minimum constriction occurring about six to eight mm behind that for /s/ and /z/. The constriction for /s/ and /z/ is made with a raised tongue tip at the alveolar ridge by MI and SC; the tongue tip raising for PK is not as distinct as evident for MI and SC while the tongue blade is raised to form the constriction for AK. For /ʃ/ and /ʒ/, the constriction for all subjects is characterized by a raised tongue blade, rather than tongue tip. The constriction has a slit-like appearance and tends to be slightly wider than that of /s/ and /z/. The A_C values, however, do not exhibit any contrasting pattern between the alveolar and the palato-alveolar fricatives. For /ʃ/ and /ʒ/, significant right-left asymmetry is visible in the constrictions' appearance for AK and SC, with more opening towards the right side. The tongue contour in the constriction region is flat or slightly convex.

Region behind the constriction: Mid-sagittal profiles for /s/ and /z/ of MI and SC reveal a marked lowering of the front of the tongue behind the constriction, with respect to the tip and back of the tongue body. The /s/ and /z/ of AK and, to a lesser extent of PK, however, are characterized by a more laminal articulation and do not reveal a significant lowering of the tongue body behind the constriction. For all subjects, the cross sectional shapes have flat to slightly convex contour in the constriction region and change significantly to concave as the posterior region of the tongue body is approached. The concave shape, which contributes to the large back cavity volume, probably results from the bracing of the tongue against the hard palate along its sides and, a relatively depressed tongue center. The degree of concavity, however, is speaker dependent. For example, subjects MI and SC reveal significant medial grooving of the front tongue body, about 2-3 cm behind the constriction for MI and 1-2.2 cm behind the constriction for SC, with the maximum medial groove depths for /s/ and /z/ reaching 11.5 mm and 10.6 mm, respectively, for MI and 8.6 mm and 10.4 mm, respectively, for SC [Figs. 2(a)-(b)]. This grooving effect contributes to the abrupt increase in the corresponding AF values behind the constriction. Subjects AK and PK, on the other hand, do not reveal noticeable medial tongue grooving. The tongue contour is almost flat for about 1 cm behind the constriction for AK, with a more (anatomical) left-favored linguo-palatal contact. For AK, the rapid increase in the AF values behind the constriction is attributed to the subject's highly domed palate, in addition to the concave tongue shape. The front tongue body region behind the constriction of PK is concave with no grooving, and the maximum medial depth reaches only 3.9 mm.

Concave cross-sectional shapes are revealed by all subjects in the posterior region, with decreasing concavity as the posterior pharyngeal wall is approached. Significant asymmetry in linguo-palatal contact is observed in PK. Raising of the back of the tongue, which also contributes to the decreased area values near the velar region, is evident in all subjects. The amount of tongue-back raising, however, is speaker-dependent, and is found to be most striking for AK. There is an artificial increase in the front region AF values due to the grooving of the curving tongue back near the VT bend. No significant differences are found between the cross-sectional tongue shapes of /s/ and /z/ for any of the subjects.

The tongue shapes for the strident fricatives suggest that apical or laminal nature of these productions are speaker-dependent. Moreover, the nature of the anterior tongue behavior is found to influence the tongue body shapes behind the constriction. The anterior tongue body behind the constriction reveals a relatively deeper grooving for the apical alveolars when compared to the laminal cases, contributing to a relatively rapid increase in AF values behind the constriction. The pressure drop due to losses at the contraction and expansion in the constriction region, on which the SPL of the turbulence source depends, is predicted to be smaller for smooth transitions when compared to more abrupt ones [3].

The tongue body behind the constriction for /ʃ/ and /ʒ/ rises slightly along its midline, before it starts sloping towards its posterior end. This results in a relatively gradual increase in the corresponding area values behind the constriction when compared to /s/ and /z/. The degree of "palatality" (relative height of the front of the tongue with respect to the back) is speaker-dependent; palatality exhibited by SC is the most striking and is reflected in a relatively rapid increase in the AF values behind the constriction. The /ʃ/ and /ʒ/ of PK are different from those of the other subjects with the tongue raising occurring at the middle rather than at the anterior tongue body; the corresponding AF's exhibit a plateau region following a small increase in values behind the constriction. It should be noted that subject PK reported, after scanning, that the tongue positions for /ʃ/ and /ʒ/ produced in a supine position inside the MRI scanner felt "backer" than those produced in normal erect posture.

The /ʃ, ʒ/ cross-sectional tongue shapes behind the constriction are quite variable across subjects. Typical groove dimensions are shown in Fig 2(c)-(d) for MI and SC, respectively. For MI and AK, a slight convex shape is evident which gradually turns concave towards the posterior region. Both PK and SC, on the other hand, do not exhibit any doming of the tongue body behind the constriction. Their tongue contours are almost flat and asymmetrical with right-favored linguo-palatal contact. Previous studies indicate that tongue asymmetries in normal sibilant productions are not uncommon [4]. It is unclear whether these asymmetries are typical or whether they are influenced by the supine position in the scanner.

Nonstrident fricatives:
The anterior tongue body for /θ/ and /ð/ is located between the teeth for all subjects. The location of the tongue tip (with respect to the teeth), however, shows significant inter-speaker variability. The tongue blade of AK and MI is between the upper and lower teeth, with the slightly upward-pointing tongue tip well in front of the incisors (5-6 mm), a production pattern prevalent in Californian dialects. Subjects SC and PK, whose backgrounds are different from those of AK and MI, on the other hand, have their tongue tips resting more or less on the lower incisors. In general, there is significant raising of the tongue body behind the anterior region, with the dorsum at a higher level than the middle and posterior part. The location and degree of this raising is speaker dependent. In addition, the tongue root retraction evident in /ð/ possibly influences the posterior tongue body raising. The overall cross-sectional tongue shapes are either flat or slightly concave. Linguo-palatal contact is evident beginning around the middle region of the tongue body. The tongue shape in the middle region shows slight concavity which gradually becomes flat as the posterior region is approached. Significant asymmetry in the tongue shapes of PK and SC is noticed in the dorsal region with the left lateral tongue body raised with respect to the right side.

The tongue body midline for /f/ and /v/ is characterized by a "bunched" position for all subjects, with a raised tongue dorsum, lowered anterior and posterior regions and a downward pointing tip. The minimum constriction occurs at the lips. The tongue tip appears at 1.5 to 2 cm behind the lip opening with distinct sublingual cavities (details in Table III). The free anterior tongue body exhibits, in general, an asymmetrical convex contour, which turns concave once linguo-palatal contact is established (at about 1.5 cm from the lip opening for AK and PK, 2.4 cm for SC and 4.5 cm for MI). Among the eight fricatives, the pattern of AF variation across the different speakers is the most different for /f/ and /v/, even though the general tongue shapes are similar. There is a trend to create a relatively large volume behind the lips by the retraction of the tongue body together with the formation of sublingual cavities. Variability in front cavity volumes may contribute to the variability in the the acoustic spectra of these sounds.

4.2 Back region: For the sake of simplicity, analysis of the back region is restricted to the uvular and pharyngeal regions. Sample back region area functions for /s/, /z/, and /f/ are shown in Fig. 1(d)-(f). The maximum measured front region area is 9.02 cm^2 (/ʒ/ of SC). Consistent AF patterns are evident although inter-speaker variabilities are more marked when compared to the front region. Among the fricatives, /ʃ/ and /ʒ/ are characterized by the largest pharyngeal areas while /f/ and /v/ have the smallest areas. Voiced sounds, in general, tend to have larger areas when compared to their unvoiced counterparts due to tongue-root retraction. For /ʃ/ and /ʒ/ the area increase is most significant in the upper pharyngeal region due to palatality. Fricatives /θ/ and /ð/ exhibit a similar pharyngeal behavior as /ʃ/ and /ʒ/. The raising of the posterior tongue body in the case of /s, z/ and /f, v/, on the other hand, contributes to the smaller AF values evident in the uvular and upper pharyngeal regions. The amount of tongue-root retraction also influences the epiglottic-vallecular volume to some extent. Increased pharyngeal volumes for the voiced fricatives may be associated with increased supraglottal pressure that accompanies voicing.

4.3 Sublingual cavities: Distinct sublingual cavities are visible in the coronal sections for the fricatives /ʃ, ʒ, f, v/ of all subjects except for /ʃ/ by AK. It is possible that AK's /ʃ/ cavity is smaller than the spatial sampling used. The sublingual areas, in general, are found to be asymmetric. Sublingual cavity volumes, obtained by 3D reconstruction of the cavities, are given in Table III. Acoustic studies have shown that the presence of sublingual cavities contributes to a relatively low frequency main spectral peak due to an increase in the effective volume of the cavity between the constriction and teeth (obstacle) [5, 6].

4.4 Length measurements: Length of VT (l_{VT}, in mm), averaged across the eight fricatives, are 168.26, 176.14, 157.7 and 175.95 for subjects AK, MI, PK and SC, respectively. Values of average l_{FR}/l_{BR} (mm) are 95.46/84.01, 101.91/93.31, 84.41/84.55, and 99.93/99.71 for subjects AK, MI, PK and SC, respectively. l_{VT} specifies the VT length measured along the midline of the vocal tract while l_{FR} and l_{BR} specify the span of the front and back regions, respectively. Lip opening values for the nonstridents are smaller than those of strident fricatives. Across speakers, no uniform contrasting behavior is observed in the lip openings of /s, ʃ, z, ʒ/.

5. SUMMARY

In this paper, a morphological analysis of the VT geometry during the production of sustained English fricatives /s, ʃ, f, θ, z, ʒ, v, ð/ obtained by MRI is presented. Inter-speaker variabilities are found to be greater in the back region than in the front region. Labio-dental fricatives exhibit the most variability across speakers. Among the strident fricatives, cross-sectional tongue shapes behind the constriction for /ʃ/ and /ʒ/ exhibit greater variabilities than /s/ and /z/. The concave tongue shapes of /s/ and /z/ result in a more abrupt area function behind the constriction when compared to that of /ʃ/ and /ʒ/. Cavity volumes are affected by both the manner of articulation and the subject's oral morphology. Between the unvoiced fricatives and their voiced counterparts, l_C values are less variable than A_C. Voiced fricatives show larger pharyngeal volumes than the unvoiced fricatives due to tongue root retraction. No gender-related differences are detected in the articulation patterns. Future work will examine the effects of the observed variability in the articulation patterns on the aerodynamics and acoustic spectra of these sounds.

Acknowledgements: The cooperation and support of the Cedars-Sinai medical imaging group is gratefully acknowledged. This work was supported in part by NSF grant IRI93-09418 and NIDCD grant DC02033.

6. REFERENCES

[1] T. Baer *et al*, "Analysis of vocal tract shape and dimensions using magnetic resonance imaging: Vowels," *JASA*, vol. 90, pp. 799–828, Aug. 1991.

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

[3] K. N. Stevens, "Airflow and turbulence noise for fricative and stop consonants: Static considerations," *JASA*, vol. 50, pp. 1180–1192, May 1971.

[4] M. Stone *et al*, "Cross-sectional tongue shapes and linguopalatal contact patterns in [s], [S] and [l]," *J. Phonetics*, vol. 20, pp. 253–270, 1992.

[5] J. S. Perkell *et al*, "Articulatory and acoustic correlates of the [s-š] distinction," in *Speech communication papers presented at the 97th meeting of the ASA*, 1976.

[6] C. H. Shadle, "The effect of geometry on source mechanisms of fricative consonants," *J. Phonetics*, vol. 19, pp. 409–424, 1991.

Table I: Location of minimum supraglottal constriction (l_C, cm) measured from lips.

Fricative	Subject			
	AK	PK	MI	SC
/ʃ/	2.6	2.08	2.78	3.37
/ʒ/	2.62	2.24	2.34	3.24
/s/	1.97	1.4	1.96	2.54
/z/	1.97	1.75	2.0	2.22
/θ/	0.95	0.89	1.01	1.51
/ð/	0.91	1.01	0.89	1.5
/f/	0.87	0.55	0.83	0.79
/v/	0.89	0.66	0.91	0.69

Table II: Area of supraglottal constriction (A_C, cm^2)

Fricative	Subject			
	AK	PK	MI	SC
/ʃ/	0.262	0.130	0.112	0.174
/ʒ/	0.299	0.098	0.124	0.252
/s/	0.146	0.098	0.142	0.296
/z/	0.117	0.136	0.159	0.247
/θ/	0.208	0.257	0.155	0.327
/ð/	0.340	0.151	0.357	0.353
/f/	0.259	0.343	0.365	0.323
/v/	0.404	0.387	0.456	0.236

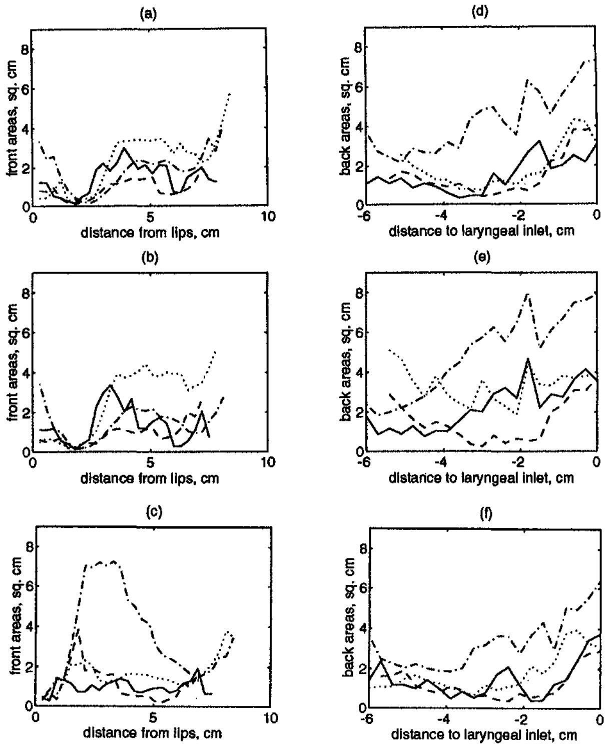

Figure 1: Sample area functions for different speakers (AK-solid, PK-dashed, MI-dot-dashed, SC-dotted). (a), (b), and (c) represent front region areas for /s/, /z/, and /f/, respectively. (d), (e), and (f) represent upper-back region areas for /s/, /z/, and /f/, respectively.

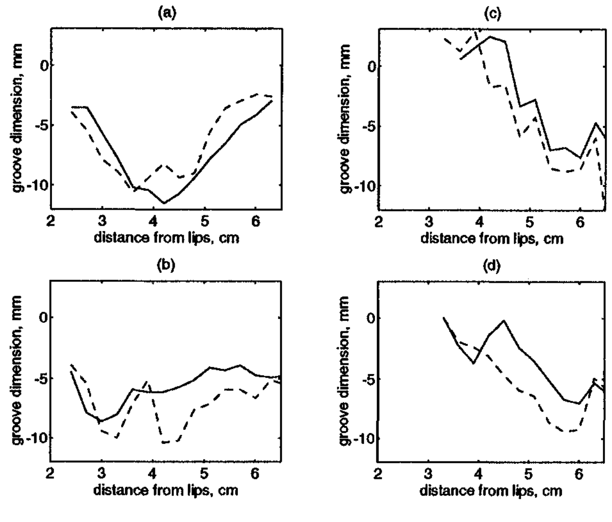

Figure 2: Medial tongue groove dimensions behind the constriction for speakers MI and SC (voiced-solid, unvoiced-dashed). (a)-(b) /s, z/ of MI and SC, respectively. (c)-(d) /ʃ, ʒ/ of MI and SC, respectively. Negative values imply grooving while positive values imply doming of the tongue with respect to its sides.

Table III: Volume of sublingual cavity (mm^3)

Fricative	Subject			
	AK	PK	MI	SC
/ʃ/	–	143.35	212.61	37.25
/ʒ/	175.42	208.02	224.35	231.86
/f/	73.79	445.37	583.77	664.00
/v/	49.35	498.95	402.09	22.65