Acoustic characteristics of the piriform fossa in models and humans

Jianwu Dang
ATR Human Information Processing Research Laboratories, 2-2 Hikaridai Seika-cho, Soraku-gun, Kyoto, 619-02 Japan

Kiyoshi Honda
ATR Human Information Processing Research Laboratories, 2-2 Hikaridai Seika-cho, Soraku-gun, Kyoto, 619-02 Japan and University of Wisconsin, 1500 Highland Avenue, Madison, Wisconsin 53705-2280

(Received 22 March 1996; accepted for publication 12 September 1996)

The piriform fossa forms the bottom part of the pharynx and acts as a pair of side branches of the vocal tract. Because of its obscure form and function, the piriform fossa has usually been neglected in the current speech production models. This study examines the geometric and acoustic characteristics of the piriform fossa by means of MRI-based mechanical modeling, *in-vivo* experiments and numerical computations. Volumetric MRI data showed that the piriform fossa is 2.1 to 2.9 cm^3 in volume and 1.6 to 2.0 cm in depth for four Japanese subjects (three males and one female). The results obtained from mechanical models showed that the piriform fossa contributes strong troughs, i.e., spectral minima, to speech spectra in a region of 4 to 5 kHz. The antiresonances were identified with increasing frequency when water was injected into the piriform fossa of human subjects in *in-vivo* experiments. Antiresonances obtained from the experiments and simulations were confirmed to be consistent with those in natural speech within 5%. Acoustic measurements and simulations showed that the influence of the piriform fossa extends to the lower vowel formants in addition to the local troughs. This global effect can be explained by the location of the fossa near the glottal end of the vocal tract.
[S0001-4966(97)04301-4]

PACS numbers: 43.70.Bk, 43.70.Aj, 43.72.Ct [AL]

INTRODUCTION

The piriform fossa, also referred to as the sinus pyriformis or Recesses Piriformis in its singular forms, consists of a pair of bilateral cavities in the hypopharynx located just above the esophageal entrance. Therefore, this space clearly belongs to the vocal tract. Except for a few previous studies (e.g., Fant, 1960; Sundberg, 1974; Fant and Båvegård, 1995), however, this structure has not been widely recognized as a functional part of the vocal tract in speech production. There seem to be at least two reasons for this neglect: one is the lack of morphological data for acoustic modeling; the other is its small size, which would appear to relegate it to minor importance as a side branch of the vocal tract.

In earlier studies, Fant (1960) made a valuable observation on the role of the piriform fossa with respect to vowel formants. He used an x-ray-based simulation to demonstrate that the acoustic effect of the piriform fossa can be seen as causing significantly lowered formant frequencies of vowels. Flach and Schwickardie (1966) reported an experimental result indicating that the piriform fossa causes sound attenuation in the frequency region above 1500 Hz. On the other hand, Mermelstein (1967) argued that theoretically the piriform fossa should not cause any major effect below 4 kHz. However, Sundberg (1974) pointed out that the piriform fossa played a significant role in forming the singing formant between 2 and 3 kHz. Lin (1990) showed that acoustic effects of the piriform fossa can vary with vowels' articulations by using a numerical model of a side branch: the effects were most prominent in $F1$ of open vowels. Fant and Båvegård (1995) found that the piriform fossa could significantly alter the density of the pole in the 3 to 5 kHz region and introduce a zero at about 5200 Hz. To obtain a realistic transfer function, they set the volume of the piriform fossa to be 50% larger than the measured volume in order to compensate for the open end correction of the piriform fossa due to the area discontinuities near the open end.

The advent of the magnetic resonance imaging (MRI) technique has made it possible to display the vocal tract, including the piriform fossa, as a three-dimensional image. A number of MRI-based studies have investigated the 3-D configuration of the vocal tract for vowels and sustained consonants (Baer *et al.*, 1991; Moore, 1992; Dang *et al.*, 1994; Narayanan *et al.*, 1995). These studies aimed at obtaining a precise area function of the vocal tract and tested the accuracy by matching the MRI-derived transfer functions to real speech spectra for the same subjects. Baer *et al.* (1991) estimated vocal tract transfer functions with and without the piriform fossa and compared the estimated formants with those from the natural utterances. They demonstrated a decrease in vowel formants due to the piriform fossa, in agreement with Fant's work, even though in their study the piriform fossa was treated not as a side branch but as an additional volume of the pharyngeal tube. Davies *et al.* (1993) used the data by Baer *et al.* (1991) to compute the transfer functions of the vocal tract with and without the

fossa. The $F1$ and $F2$ of the vowel /a/ decreased by about 5% when the fossa was incorporated in the vocal tract as a side branch. The above studies suggest that the piriform fossa does cause significant effects on speech spectra; furthermore, the studies imply that estimation of vocal tract transfer functions from the area function could be erroneous if consideration is not given to the piriform fossa. In accordance with this view, many studies have shown that an MRI-derived transfer function tends to exhibit slightly higher formant frequencies than those from real speech. Yang *et al.* (1994) explained the discrepancy between estimated and measured formants by underestimation of the vocal tract length in the MRI data. Judging from the earlier studies, however, the piriform fossa might be a critical factor in accounting for the discrepancy.

The above studies have provided conclusions resulting in both agreements and disagreements with respect to the detailed effect of the piriform fossa, and the exact role of this structure in natural speech is not yet very clear. The present study aims to explore the acoustical characteristics of the piriform fossa by means of geometric and acoustic measurements. Volumetric MRI data were obtained for four subjects. Mechanical models were constructed to replicate a partial vocal tract from the glottis to the bend of the velum based on the volumetric data from one of the subjects. Those models were used to examine the acoustical effects of the piriform fossa on the radiated sounds by artificially changing the air volume of the cavities. A similar approach was repeated on human subjects by injecting water into the fossa during sustained vowel production. Antiresonances derived from the geometric data were compared with those from natural speech produced by the subjects.

I. MORPHOLOGICAL MEASUREMENT OF THE PIRIFORM FOSSA

Determining the acoustic characteristics of the piriform fossa requires detailed morphological information. In this study, we made volumetric MRI measurements to obtain the dimensions of the piriform fossa and then developed mechanical and numerical models based on the results of the volumetric analysis.

A. Method for MRI measurement of the piriform fossa

Volumetric MRI data of the vocal tract were collected from four Japanese subjects (three males and one female) by the standard spin echo method. The relaxation time (TR) was 800 ms and the excitation time (TE) was 18 ms. A 25 cm×25 cm field of view for a slice of an image was digitally represented by a 256×256 pixel matrix (see Dang *et al.*, 1994). The MRI data consisted of 26 slices for the transverse plane and 24 slices for the coronal plane. The slice thickness was 0.5 cm for both orientations. Each slice was resampled into an image of 250×250 pixels so that each pixel represented a 0.1×0.1 cm square. The resampled slices were further processed to form volumetric data consisting of 0.1 cm^3 voxels by means of interslice image interpolation. The procedure was performed with a commercial software (VoxelView) on a workstation (IRIS Indigo).

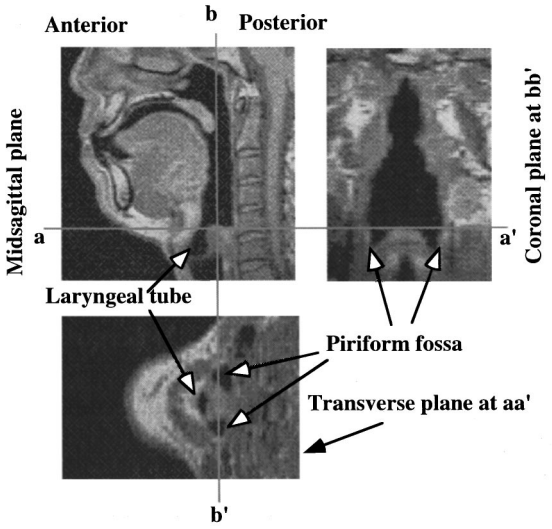

FIG. 1. A three-plane image of the volumetric MRI data of the vocal tract. The coronal and transverse planes show the gross shape of the piriform fossa. Line aa' indicates the arytenoid apex plane.

B. Three-dimensional shape of the piriform fossa

Figure 1 shows a three-plane image of the volumetric MRI data of the vocal tract. The coronal and transverse planes show that the piriform fossa consists of two bilateral cavities located on either side of the laryngeal tube. The transverse section indicates that each cavity opens to the pharynx at the horizontal plane passing through the apex of the arytenoid prominence, which is marked by the line aa' in the figure.

Figure 2 shows a sketch of the vocal tract in the vicinity of the piriform fossa. The bottom of the vocal tract divides into three small branches: the laryngeal tube (vestibule of the larynx) and the bilateral cavities of the piriform fossa. The laryngeal tube is a short conduit of about 2 cm long for our

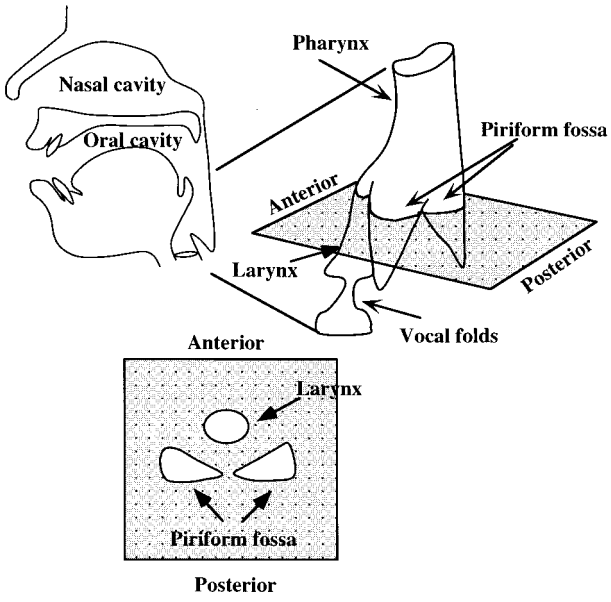

FIG. 2. A diagram of the vocal tract shape in the vicinity of the piriform fossa. The horizontal plane shows a section near the open end of the piriform fossa.

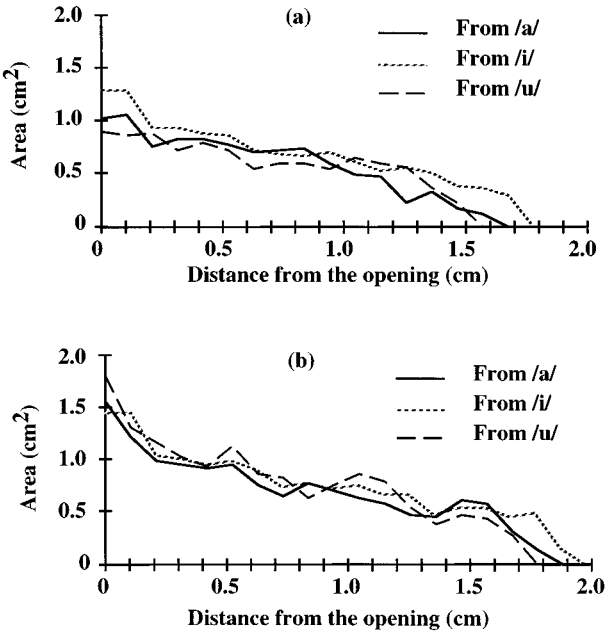

FIG. 3. Area functions of the piriform fossa based on the MRI data from subject KH. The areas are measured from the arytenoid apex plane to the bottom of the piriform fossa: (a) the left cavity and (b) the right cavity.

male subjects, bounded by the vocal folds at the bottom and by the aryepiglottic fold at the top. The piriform fossa is located behind the laryngeal tube, and forms the bottom of the pharynx. The cavities of the fossa open into the main pharynx, sharing the aryepiglottic folds as a common boundary with the laryngeal tube. However, the opening of the piriform fossa is not demarcated by a clear border because the walls of the piriform fossa continue to the vallecula medially and to the pharyngeal walls laterally and posteriorly. With reference to the direction of sound propagation, the piriform fossa comprises a pair of side branches bifurcating from the vocal tract. Since the shape of the cavities is relatively stable during phonation, the piriform fossa is expected to contribute relatively invariant antiresonances to the transmission characteristics of the vocal tract.

Dimensions of the left and right cavities of the fossa were measured for the Japanese vowels /a/, /i/, and /u/ based mainly on the transverse data with reference to the coronal data. Depth of the fossa (the distance from the bottom to the arytenoid apex plane) was determined using the coronal data. Cross-sectional shapes of the fossa were measured on the transverse slices for the entire extent from the bottom of the piriform fossa to the arytenoid apex. The results are shown as area functions in Fig. 3 for the left and right cavities of subject KH. The depth and volume of the piriform fossa are shown in Table I for the four subjects. The depth values over the three vowels range from 1.6 to 2.0 cm for all subjects. The typical value is about 1.8 cm. HH showed a relatively larger variation in depth across the vowels than the other subjects. The volume ranges from 1.0 to 1.4 cm^3 for each side of the cavities. The female subject showed a smaller cavity volume than the males. Within a subject, the variation of the volume across vowels is the smallest for RY and the largest for SM. Generally, the morphological data showed no significant asymmetry between the left and right sides.

TABLE I. Depth (D) and volume (V) of the left and right cavities of the piriform fossa obtained from volumetric MRI data for four subjects.

Subject	Vowel	Vr(cm^3)	Vl(cm^3)	Dr(cm)	Dl(cm)
HH(M)	/a/	1.401	1.346	1.7	1.6
	/i/	1.027	0.977	2	1.7
	/u/	1.248	1.167	2	2
KH(M)	/a/	1.479	1.287	1.9	1.8
	/i/	1.481	1.437	1.9	1.9
	/u/	1.345	1.065	1.8	1.7
SM(M)	/a/	1.147	1.112	1.7	1.8
	/i/	1.621	1.77	1.9	2
	/u/	1.464	1.49	1.7	1.8
RY(F)	/a/	1.159	0.921	1.7	1.8
	/i/	1.08	0.953	1.7	1.7
	/u/	1.043	0.997	1.8	1.7

C. Acoustically effective cavities of the piriform fossa

The MRI data show that the piriform fossa extends slightly above the arytenoid apex plane and continues to the aryepiglottic folds in /a/ or further to the lateral glossoepiglottic folds in /i/ and /u/, depending on the position of the epiglottis. Therefore, the extended portion above the arytenoid apex plane shapes the acoustically effective opening end of the piriform fossa. Figure 4 illustrates the coronal MRI image of the piriform fossa and the schematized cavity shape by a cone and cylinder model. Here, the piriform fossa is modeled by two cascaded portions: the lower ''cone'' portion from the bottom of the fossa to the arytenoid apex plane; and the upper ''cylinder'' portion of the oblique output end above the plane. In the figure, point A is the intersection of the effective open end and the arytenoid apex plane, and point B is the intersection of the open end and the lateral wall of the pharynx. L_B represents the length from the arytenoid apex plane to point B. According to the MRI data, L_B can be roughly evaluated by an empirical formula:

$$L_B = 1.3/D, \tag{1}$$

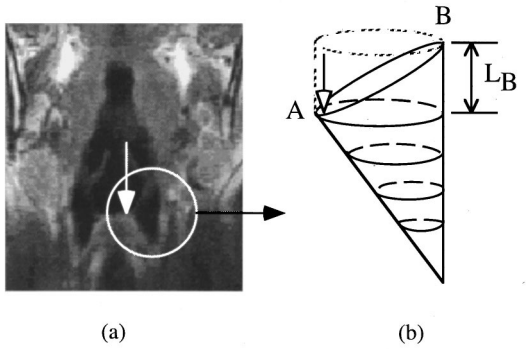

FIG. 4. A coronal MRI slice of the piriform fossa (a) and a diagram of the effective cavity of the piriform fossa (b). (White arrows indicate the arytenoid apex plane; L_B is length of the cylinder portion above the plane.)

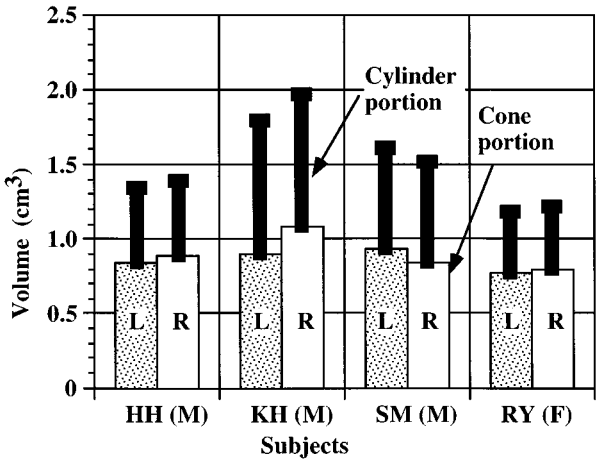

FIG. 5. Volumes of the lower ''cone'' and upper ''cylinder'' portions of the piriform fossa for four subjects (L: left cavity, R: right cavity).

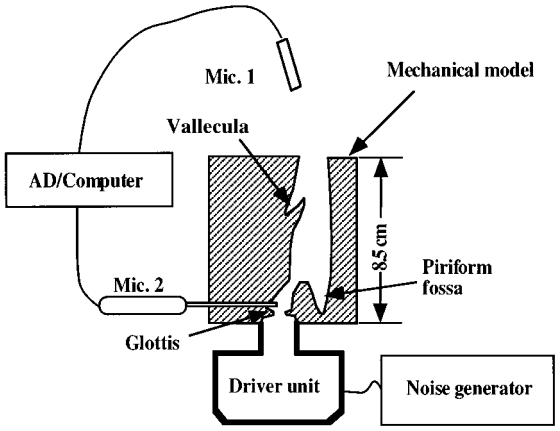

FIG. 6. Experimental setup for measuring the acoustic effects of the piriform fossa using a mechanical model constructed from volumetric MRI data.

where D denotes the diameter of the cross-sectional area of the piriform fossa in the arytenoid apex plane and the area is assumed to be circular.

Using Eq. (1), the effective cavity of the piriform fossa was calculated for the four subjects. Figure 5 shows the volumes of the cylinder and cone positions for the left and right cavities, which are averaged over three vowels /a/, /i/, and /u/. The ratio of the cylinder to the cone portions ranged from 0.4 to 0.8 for the four subjects. This result supports the suggestion of Fant and Båvegård (1995), who used an estimated volume in their study that was about 50% larger than the measured one (i.e., the cone portion).

II. ACOUSTIC MEASUREMENT USING THE MECHANICAL MODELS

The volumetric MRI data allow us to examine the acoustic characteristics of the piriform fossa by either mechanical or numerical models. The numerical model requires parameters to define the acoustic properties of the geometry, such as an open end correction for the cavities of the piriform fossa. However, the estimation of accurate parameter values is difficult because of the complex shape of the structure. For this reason, a mechanical model of the vocal tract was first used in this study. The coefficient of the open end correction is derived from the experiments on the mechanical models and used in numerical simulations in the latter sections. The mechanical models also served in a pilot experiment of injecting water into the piriform fossa of the human subjects.

A. Experimental procedure

Three mechanical models of the vocal tract for the Japanese vowels /a/, /i/, and /u/ were constructed based on the MRI data obtained from subject KH to be used for acoustic experiments. The air-tissue pattern in the transverse MRI image was carved out of a vinyl chloride plate with a size of 7 cm×7 cm×0.1 cm. The vertical part of the vocal tract was replicated by the carved plates. The models were about 8.5 cm long, from about 1 cm below the glottis to the vocal tract bend near the velum. The openings of the piriform fossa in the models were about 3.5 cm from the input end (glottal side) and about 5 cm from the radiating end (velar side). The vallecula was at about 6 cm from the input end in the models for vowels /i/ and /u/, while it was collapsed in vowel /a/. Hereafter, the mechanical model is referred to as the M model and the numerical model used in the later sections is called the N model. The M models are labeled M model /a/, M model /i/, and M model /u/, corresponding to their original vocal tract shapes, respectively.

The acoustic effects of the piriform fossa on the transfer function of the M models were examined by an experiment using the two-point sound pressure method (Dang and Honda, 1996a). The sound pressures were recorded at an inside point near the glottis and at the radiating end, while the model was excited by an external source sound. The transfer function obtained from the two-point pressure is known to exactly reflect the acoustic properties of the tube segment between the two points, since the effects of the sound source and the geometric shape upstream from the inside point are excluded from the estimation. Figure 6 shows the experimental setup for measuring the transfer function of the M models. The sound pressures inside and outside the M model were recorded by two microphones. Microphone M1, a B&K 4003, was placed about 6 cm away from the radiating end. Probe microphone M2, a B&K 4182, was used to record the pressure inside the model just above the glottis via a flexible probe tube, which was inserted through a hole in the lateral wall of the M model. The probe tube with a matched impedance to the microphone was 5 cm long, 0.165 cm in outer diameter and 0.076 cm in inner diameter. A white-noise signal was generated by an FG-143 function generator (NF Circuit Design Block Co.) and fed into the M model using an SG-505FRP horn driver unit (Goto Unit Co.). The joint between the M model and the throat of the horn driver was sealed with plasticine to prevent sound leakage.

The experiment on the M models was conducted in an anechoic room under four conditions: both cavities filled with plasticine, either the right or left cavity filled, and both cavities empty. Two pressure signals were sampled at a rate

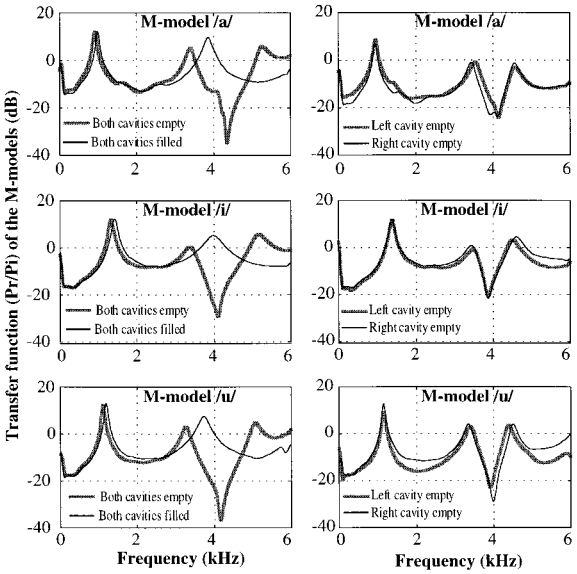

FIG. 7. Pressure-to-pressure transfer function (Pr/Pi) of the mechanical models when one or both cavities are empty, or filled with plasticine (Pr: radiated pressure; Pi: intrapressure).

of 44.1 kHz. A room temperature of 20 °C was maintained during the measurements, and the sound velocity is expected to be 34 300 cm/s. The outside sound pressure near the joint was about 25 dB lower than that at the radiating end, which implies that sound leakage from the joint was negligible in this measurement. An FFT-derived cepstrum analysis (Imai and Abe, 1979) was applied to obtain spectral envelopes of recorded signals. The cepstrum coefficients were weighted by a Hamming window of 0.05 second for white noise and 1.2 times the fundamental period for speech sound.

B. Changes in the transfer function by varying the models' cavity volume

Transfer functions of the three M models were estimated by the two-point sound-pressure method under the four conditions described above. The results are shown in Fig. 7 for the M models (left) with and without the both cavities and (right) with either the left or right cavity alone. Note that though the models represent only a part of the vocal tract, they exhibit the same antiresonances on the transfer function as they do in the whole vocal tract as far as the side branches in the models are concerned.

In Fig. 7(left), the thick lines show the spectra obtained under natural conditions, i.e., with both cavities empty; the thin lines indicate the results measured when the two cavities were filled with plasticine. Deep troughs are seen at about 4 kHz in the spectra obtained under natural conditions for all the M models of /a/, /i/, and /u/. When the cavities were filled bilaterally, the troughs were replaced by peaks, resulting in an increase of more than 30 dB in spectral level near 4 kHz. The large difference in the spectra indicates that the piriform fossa behaves as a side branch of the vocal tract and causes a significant effect on the transfer function. In Fig. 7(right), the thick lines show the spectra obtained from the M models with the left cavity empty, and the thin lines represent the

TABLE II. Antiresonance frequencies caused by the empty cavities of the piriform fossa in the transfer function of the mechanical models (Hz).

M models	Both cavities	Left cavity	Right cavity
M model /a/	4359	4148	4100
M model /i/	4054	3855	3867
M model /u/	4160	3925	3996

results with the right cavity empty. The figure shows almost identical spectral shapes for the left and right cavities except for a small discrepancy in the troughs' shape. This implies that the left and the right cavities are acoustically symmetrical in this subject. This observation is consistent with the morphological observation discussed in Sec. I.

The frequencies of these troughs were measured for each condition and are listed in Table II as the antiresonance frequencies of the models' piriform fossa. The antiresonance frequencies ranged from 4050 to 4360 Hz in the case of both cavities empty. The M model /a/ shows a slightly higher antiresonance frequency than the others. In all three M models, the antiresonance frequency caused by a single cavity was lower than that caused by both cavities together. This phenomenon can be explained by a larger open end effect of the cavity because the area ratio of the piriform fossa to the pharyngeal cavity at their boundary is larger in the case of a single cavity than it is under natural conditions. This implies that an appropriate open end correction is necessary for incorporating the piriform fossa into the transmission line model. In this regard, measurements and numerical computations have been conducted to obtain the optimum value of the end correction coefficient (Dang and Honda, 1996b). The measurements were carried out on the M models while the models' fossa was filled with plasticine from the bottom (0 cm) up to a level of 0.9 cm in 0.1-cm intervals. In the computations, the piriform fossa was schematically represented by the cone and cylinder portions as shown in Fig. 4. The length of the cylinder portion of the fossa (L_B) was derived by Eq. (1). The antiresonance frequency was then computed for each M model. The optimum value of the open end correction coefficient was estimated by matching the computed antiresonances to the measured ones. The value of 0.75 met the given condition that the computed antiresonances were consistent with those measurements within 4% for the three M models.

Figure 7 also shows a global acoustic effect of the piriform fossa in addition to the local troughs near 4 kHz. In the figure, a relatively small but non-negligible difference in the spectral peaks can be seen in the lower frequency region below 3 kHz. Peak frequencies of the model's transfer function are listed in Table III. As a general tendency, the piri-

TABLE III. Frequencies of the first two peaks in the transfer function (Pr/Pi) of the mechanical models (Hz).

	Empty cavities		Filled cavities	
M models	$P1$	$P2$	$P1$	$P2$
M model /a/	925	3398	1007	3855
M model /i/	1335	3348	1429	3978
M model /u/	1113	3246	1207	3714

form fossa lowers the frequencies of the peaks that are in the wide frequency region below the antiresonance of the fossa. The changes in the frequency of the first peak are 82 Hz for /a/, 94 Hz for /i/, and 94 Hz for /u/. For the second peak, they are 457 Hz for /a/, 630 Hz for /i/, and 468 Hz for /u/. The results show that the piriform fossa lowers the frequency of the spectral peaks about 8% in the region near 1 kHz and about 13% in the region above 3 kHz. Note that since the M models represent only a part of the vocal tract, the peaks do not correspond to true vowel formants. However, the observed frequency shifts should be representative of those of true formants that are located in the frequency region.

C. Changes in antiresonances of the piriform fossa with water injection

The direct approach used to explore the acoustic effects of the model's piriform fossa is not applicable to living human subjects. However, a similar experiment could be performed by injecting water into the fossa of human subjects during sustained phonation. Under these conditions, the effects of the piriform fossa could be viewed as increasing antiresonance frequency with a decrease in the air volume of the cavities. This hypothesis was first tested by a pilot experiment on the mechanical models.

The experimental setup for the pilot experiment was based on Fig. 6, where probe microphone M2 was removed and a membrane of polyvinylidene chloride was placed between the M model and the neck of the driver unit to prevent water from leaking into the driver unit. A piece of rubber foam was placed into the model's glottis to reduce acoustic influence of the air volume change in the piriform fossa on the driving impedance of the model. The joint between the M model and the driver unit was sealed with plasticine. The amplitude of the excitation sound source was kept constant during the measurement. Radiated sound from the model was recorded by microphone M1 while water was injected into the piriform fossa at a constant rate through a thin flexible tube.

Figure 8 illustrates running spectra of the recorded sound from M model /a/. Each curve in this figure corresponds to a frame of 46-ms duration, and the frame-to-frame interval was about 130 ms. Since the excitation source was constant during the measurements, the changes in the spectra are considered to be caused by the changes in the air volume of the fossa only. In the initial condition with no water injected, the spectral curves show a constant trough at about 4300 Hz. After about 1.2 s, water is injected into the right cavity of the piriform fossa. The trough moves toward a higher frequency as the air volume of the right cavity decreases due to injected water. Then, another trough by the left cavity is exposed and remains at about 4100 Hz. The trough caused by the right cavity disappears from the spectra as the cavity is completely filled with water after about 4 s. Furthermore, the trough associated with the left cavity begins to increase in frequency as the water flows into the left cavity. Unlike the gradual motion for the right cavity's trough, the one caused by the left cavity moves rapidly. The cause of this rapid rise of the trough was confirmed by a visual observation of water flow in the fossa: the water reserved in the right cavity by surface tension suddenly flows into the left cavity through the postarytenoid space that interconnects the two cavities. Comparing the frequencies of the peaks before and after the water injection, it can be seen that the first peak increases about 9% when the piriform fossa was completely filled with water.

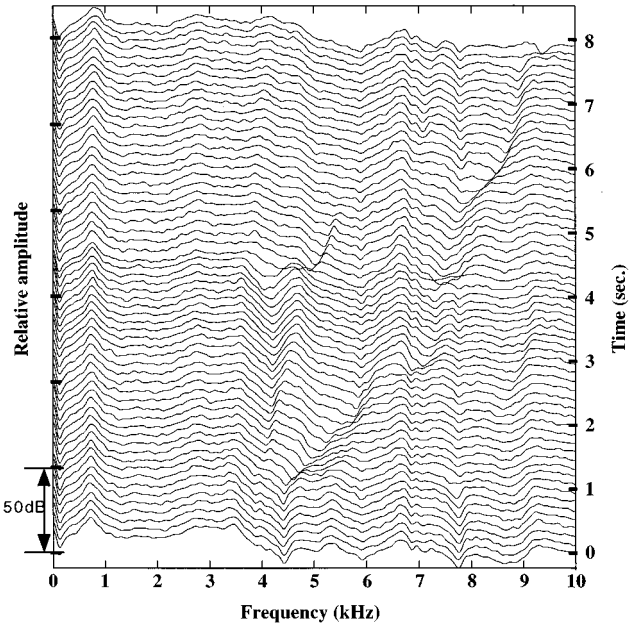

FIG. 8. Running spectra of the radiated sound from M model /a/ excited by constant sound source. Motion of the troughs is shown while water is injected into the piriform fossa.

III. ANTIRESONANCES OF THE PIRIFORM FOSSA IN SPEECH SPECTRA

The above results indicate that water injection can be used to examine the acoustic effects of the piriform fossa in humans. Water has a density close to that of muscles, and is an adequate material to fill the piriform fossa of humans if the injection is conducted carefully. In this section, the antiresonances of the piriform fossa are examined by a water injection experiment and using natural vowel utterances.

A. Water injection experiment on human subjects

The same procedure of water injection was applied to two male subjects JD and KH. A flexible tube with a 0.3-cm o. d. and a 0.2-cm i.d. was inserted along the nasal floor and passed through the nasopharyngeal port toward the piriform fossa. The tip of the flexible tube was placed above the right cavity under fiberoptic video-endoscopy. Warm water, at about body temperature, was injected into the right cavity through the flexible tube. When the right side was filled, the water was seen to flow into the left cavity. Radiated oral sounds of sustained vowels /a/, /i/, and /u/ were recorded during the water injection.

Figure 9 shows running spectra of the vowel /a/ for JD. The motion pattern of the troughs shown in this figure is similar to that seen in the M models. In the figure, a trough appears at about 5200 Hz before the water was injected. After about 0.2 s, the trough caused by the right cavity rises

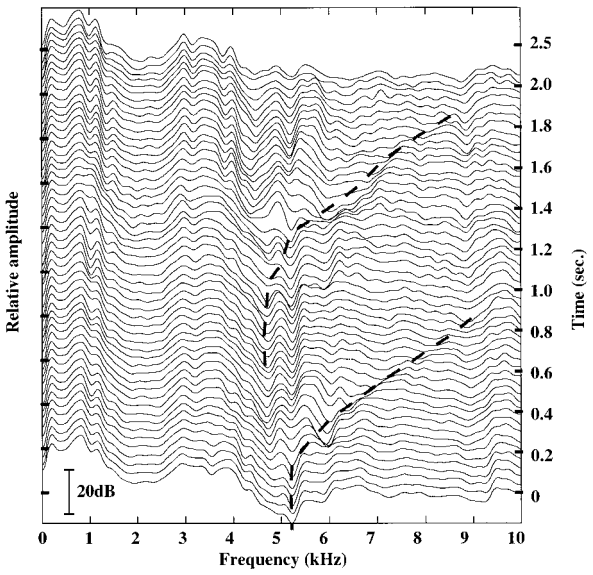

FIG. 9. Running spectra for sustained vowel /a/ produced by subject JD while water is injected into the piriform fossa.

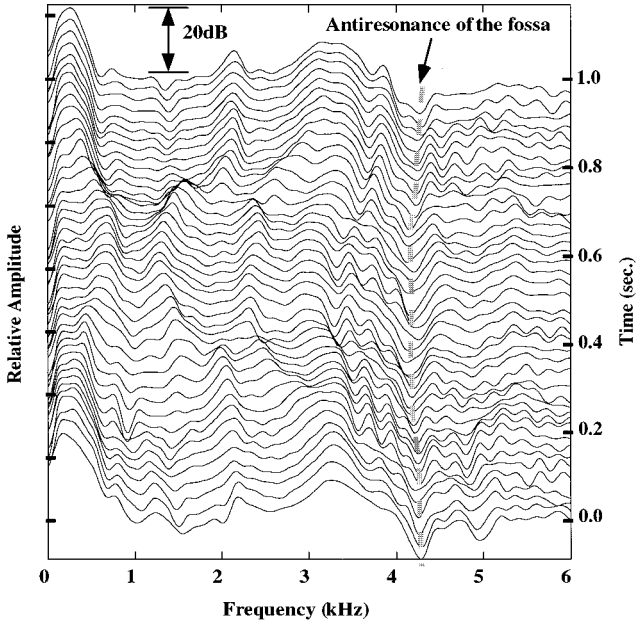

FIG. 10. Running spectra of /iai/ sequence produced by subject KH in the supine position.

as the cavity is filled with water. This trough disappears from the spectra when the right cavity is full, which is observed at about 1 s. The trough for the left cavity begins to rise as the water flows into the left side, which occurs approximately after 1 s, and disappears when both the right and left cavities become full after about 2 s. The running spectra indicate that the piriform fossa causes two antiresonances around 5 kHz in speech spectra for this subject, although the two antiresonances usually appear as one trough in speech spectra. The same trough pattern in the antiresonances was also confirmed in the speech spectra for KH.

Although both the experiments on the M models and on humans demonstrated a consistent trough pattern, the expected changes in the lower formants were not obvious in humans. In particular, $F1$ remained almost constant during water injection. There seem to be at least two possible causes for the stability of $F1$. First, a physiological reflex to the injected water could result in a small change in articulatory posture for the vowels. Second, the subjects might adjust articulator positions to compensate for the acoustic effect of decreasing air volume by means of auditory feedback. An additional experiment was performed to verify the latter possibility (i.e., auditory feedback) by applying a loud masking noise during water injection. The result did not show the expected changes in the first formant, but rather supported the former explanation (i.e., physiological reflex to water injection).

B. Antiresonance of the piriform fossa in natural speech

It was shown in the above experiments that the piriform fossa causes a trough in the 4 to 5 kHz region. A question is raised as to whether such a trough can be found in natural speech. To answer this question, natural vowel utterances were analyzed for the four subjects who served for the MRI experiment. Speech materials consisted of isolated vowels /a/, /i/, /u/, and vowel sequences (/aia/, /aua/, /iai/, /iui/, /uau/, and /uiu/). Sound recording was carried out in an anechoic room at a sampling rate of 44.1 kHz, where the subjects repeated the speech materials twice at a natural speech rate in two body postures: upright and supine. Figure 10 shows running spectra of /iai/ recorded in the supine position for KH. While the formants in this figure show articulatory changes from /i/ to /a/ to /i/, a trough at about 4300 Hz remained almost constant during the vowel sequence. Referring to the results from human and model experiments, it is reasonable to judge that the trough is the antiresonance caused by the piriform fossa.

Speech samples recorded from the four subjects in the supine position were analyzed for the stable segments in the three Japanese vowels. Figure 11 shows the means and standard deviation of the antiresonance frequencies for the four subjects. The coefficient of variation (CV, the ratio of the

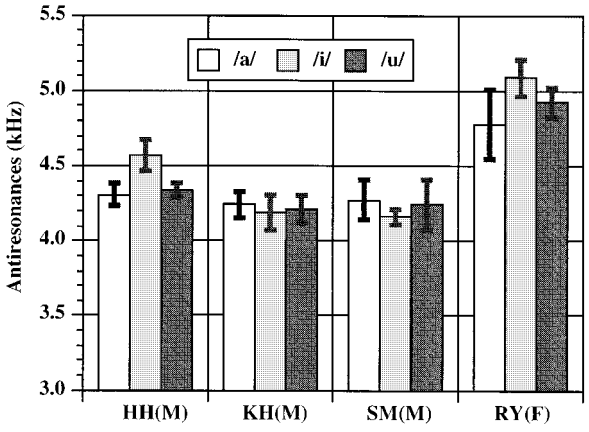

FIG. 11. Means and standard deviations of antiresonance frequencies caused by the piriform fossa measured for each vowel from vowel sequence data of four subjects.

standard deviation to the means) for each vowel was about 5% for all subjects, except that of /a/ for subject RY, which was about 10%. The CVs among the male subjects were within 5%, whereas it was about 15% between the female and the males. The female subject showed higher antiresonance frequencies than the males. The results indicate that antiresonances caused by the piriform fossa appear constantly in the spectra of the natural vowel sequences.

IV. SIMULATIONS USING NUMERICAL MODELS

The MRI data and M-model experiments in this study provide a clear view of the geometric and acoustic characteristics of the piriform fossa. These findings were adopted to design a numerical model (N model) of the vocal tract based on a transmission line model. The performance of the N model was tested in a comparison of simulation results and real speech spectra. The local and global acoustic effects of the piriform fossa were examined by model simulation and discussed on the basis of acoustic theories.

A. Comparison of the antiresonances from the N model and natural speech

The numerical model is a transmission line model designed according to the morphological data of our subjects. The basic design of the model is a pair of side branches attached to the main tract at 2 cm above the glottis. The cone and cylinder model of the piriform fossa as shown in Fig. 4 is represented by area functions with section lengths of 0.1 cm. The open end correction coefficient of the fossa was 0.75, as discussed in Sec. II. The radiation impedance of the vocal tract was approximated by a cascade concatenation of radiation resistance and radiation inductance (Caussé *et al.*, 1984), which is valid for a wide frequency region of $kr<1.5$, where k is the wave number and r is the radius of the radiating end.

The accuracy of the N model was tested by a comparison of the computed and measured transfer functions of the M models. The N model computations were confirmed to be consistent with the measured M model spectra within 2% for both resonances and antiresonances below 6 kHz. Then, the antiresonance of the piriform fossa was calculated using N models of the entire vocal tract for the four subjects. They are shown in Fig. 12, along with those obtained from natural speech for each subject. The antiresonance frequencies predicted by the N models were consistent with those from the speech data within 5%, except for a vowel /i/ of the subject HH. The result shows that the model gives a realistic description of the antiresonances of the piriform fossa. Reviewing the results above, the antiresonance frequency (F) of the piriform fossa and the depth (D) of the cone portion approximately satisfy the relation $F=c/4D$ after taking into account the cylinder portion, where c denotes the sound velocity.

B. The global and local effects of the piriform fossa

The global effect of the piriform fossa on the lower formants has been demonstrated by earlier studies (Fant, 1960; Baer *et al.*, 1991; Davies *et al.*, 1993; Fant and Båvegård, 1995). The purpose of the numerical simulation in this section is to examine the causality of the global effect of the piriform fossa. The result of simulation is discussed in terms of the plausibility of two possible treatments of the piriform fossa in vocal tract modeling: as side branches or as an additional volume.

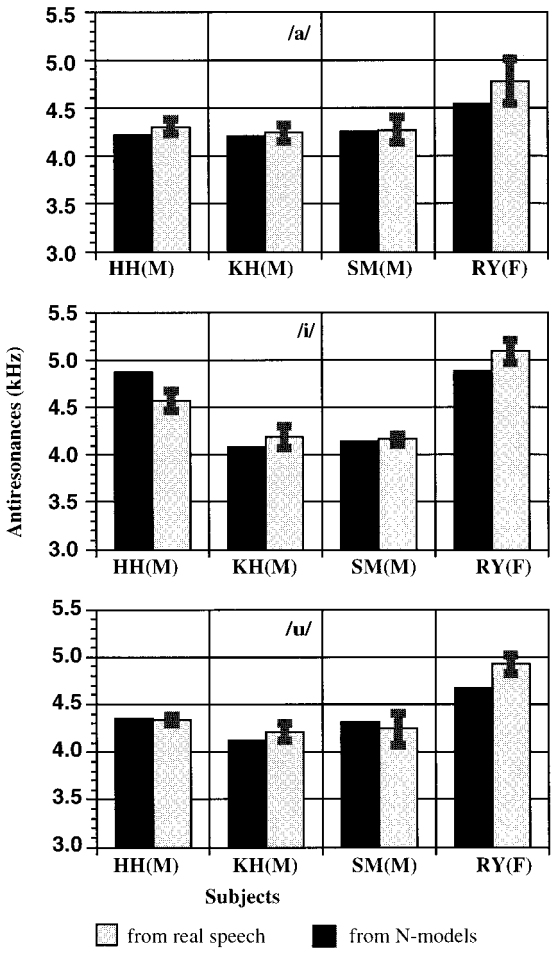

FIG. 12. Comparison of the antiresonance frequencies derived from N models and real speech for four subjects.

N model simulations were carried out for the vowels /a/ and /i/ under three conditions in which the piriform fossa was treated as a side branch (SB), as an additional volume (AV), and with no fossa (NF) as a control condition. The vocal tract configuration of the N model was based on volumetric MRI data of KH. Real speech spectra from the subject were obtained by homomorphic analysis (Markel and Gray, 1976). Figure 13 shows the speech spectra and computed velocity-to-velocity transfer functions from the glottis to the lips of the N models. In the case of /a/, $F1$ decreases by 9% in the AV treatment, and by 10% in SB treatment in comparison with the NF condition. In contrast, the effects on $F2$ and $F3$ of /a/ are relatively small. The differences in $F2$ and $F3$ were 2% and 3% in SB, while they were not seen in AV. In the case of /i/, the first two formants were lowered 3% in both AV and SB for $F1$, and 2% in AV and 6% in SB for $F2$, respectively. In comparison with the condition with-

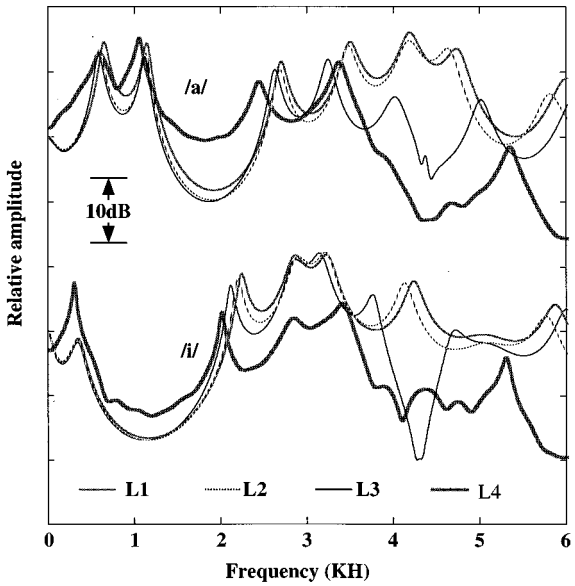

FIG. 13. Velocity-to-velocity transfer function from the glottis to the lips computed using N models, whose area functions were obtained from KH for vowels /a/ and /i/ under conditions: without the piriform fossa (L1), treating the fossa as an additional volume (L2) and as a side branch (L3). L4 shows speech spectra from the same subject.

out the piriform fossa, both AV and SB treatments demonstrated a global effect on the lower formants. The extent of the changes in the lower formants is higher than the different limens shown by Kewley-Port and Watson (1994) and Hawks (1994), who pointed out that formant-frequency discrimination is 14 Hz for $F1$ (<800 Hz), and 1.5% for $F2$. Therefore, it can be expected that the global effect causes some perceptual difference. This expected effect was confirmed in our informal listening test.

A few theoretical discussions are found in the literature regarding the acoustic effects of local volume change in the vocal tract. According to the perturbation theory (Schroeder, 1967), the effect of an additional volume of the vocal tract on formant frequencies is approximately proportional to the volume increment and the square of the sound pressure at the location of the additional volume. The same phenomenon can also be explained by the transmission line theory. The effect of a side branch on vowel formants is proportional to the admittance of the branch and the square of the sound pressure at the branch location in the vocal tract. Both theories agree in their view that the manner of formant frequency change depends on the sound pressure at the location where a small volume change takes place. At the closed end of the vocal tract sound pressure is near a maximum for all formant frequencies. Since the piriform fossa is located near the closed end of the vocal tract, it plays an important role for almost all formants. In the region much lower than the antiresonance frequency of the piriform fossa, the distribution of sound pressure in the vocal tract is expected to be the same in AV and SB treatments. Accordingly, the global effect of the fossa on the lower formants can be represented equally well regardless of whether the fossa is treated as a side branch or as an additional volume. However, the AV treatment demonstrates an incorrect behavior in the frequency region close to the antiresonance frequency of a branch because the perturbation theory cannot predict the pressure disturbance caused by the antiresonance of the branch.

Summarizing our experimental data and the above theoretical accounts, SB treatment demonstrates much more realistic performance than AV treatment in the region above 2 kHz, though they both can represent real speech spectra relatively well in the region below 2 kHz. Taken together the local and global features of the acoustic effects, it is reasonable to conclude that the piriform fossa should be modeled as side branches of the vocal tract.

V. CONCLUSIONS

This study examined the acoustic effects of the piriform fossa on the vocal tract transfer function by conducting both human experiments and model simulations. Our results indicate that the piriform fossa forms an important part of the vocal tract in realizing natural speech spectra. Furthermore, the results suggest that the function of the piriform fossa must be incorporated in any realistic model of speech production.

The morphology of the piriform fossa was investigated using volumetric MRI images for four subjects. Mechanical models were constructed from the volumetric data of one of the subjects and were used for acoustic investigation. The effective cavity of the modeled piriform fossa consists of two cascaded portions: the cone portion from the bottom of the fossa to the arytenoid apex plane and the cylinder portion above the plane. The ratio of the cylinder portion to the cone portion ranged from 0.4 to 0.8 for the four subjects.

The acoustic characteristics of the piriform fossa were investigated on the mechanical models by manipulating the air volume of the fossa. The results showed that the fossa causes troughs in the transfer function of the vocal tract in the frequency region between 4 to 5 kHz. The fossa not only affects spectral shape in the vicinity of its antiresonance, but also decreases resonance frequencies in the lower frequency region. Observations of an *in-vivo* experiment on human subjects and in natural vowel sequences showed that the antiresonance of the fossa constantly appears in natural speech utterances as well as in sustained vowels.

Numerical models were designed to simulate the acoustic behaviors of the piriform fossa. Computed antiresonances of the piriform fossa were consistent with those obtained from natural speech, generally within 5% for the four subjects. The numerical model demonstrated that the piriform fossa causes global effects on the lower formants because it is located near the closed end of the vocal tract. Considering both global and local effects, the piriform fossa should be treated as a side branch attached near the glottal end of the vocal tract rather than as an additional volume of the pharynx.

ACKNOWLEDGMENTS

The authors would like to express their appreciation to Christine Shadle for her helpful discussions. The authors also would like to thank Gunner Fant and Osamu Fujimura for

their instructive comments. We also thank Naoki Kusakawa and Hiroyuki Hirai for their assistance in the experiments.

Baer, T., Gore, J., Gracco, L. C., and Nye, P. W. (**1991**). ''Analysis of vocal tract shape and dimensions using magnetic resonance imaging: vowels,'' J. Acoust. Soc. Am. **90,** 799–828.

Caussé, R., Kergomard, J., and Lurton, X. (**1984**). ''Input impedance of brass musical instruments—comparison between experiment and numerical model,'' J. Acoust. Soc. Am. **75,** 241–254.

Dang, J., Honda, K., and Suzuki, H. (**1994**). ''Morphological and acoustical analysis of the nasal and the paranasal cavities,'' J. Acoust. Soc. Am. **96,** 2088–2100.

Dang, J., and Honda, K. (**1996a**). ''A new method for measuring antiresonances of the vocal tract transmission characteristics: an experiment study of acoustic tubes,'' J. Acoust. Soc. Jpn. (E) **13**, 93–99.

Dang, J., and Honda, K. (**1996b**). ''Acoustical modeling of the vocal tract based on morphological reality: Incorporation of the paranasal sinuses and the piriform fossa,'' Proc. of 4th Speech Production Seminar, 49–52, Grenoble.

Davies, L., McGowan, R., and Shadle, C. (**1993**). ''Practical flow duct acoustics applied to the vocal tract,'' in *Vocal Fold Physiology: Frontiers in Basic Science*, edited by I. Titze, pp. 93–142.

Fant, G. (**1960**). *Acoustic Theory of Speech Production* (Mouton, The Hague), 2nd ed., 1970.

Fant, G., and Båvegård, M. (**1995**). ''Parametric model of VT area functions: Vowels and consonants,'' EU report, Speechmaps (Esprit/BR 6975) Deliv. 28 WP2.2.

Flach, M., and Schwickardie, H. (**1966**). ''Die Recessus Piriformes unter phoniatrischer Sicht,'' Fol. Phoniat. **18**, 153–167.

Hawks, F. W. (**1994**). ''Difference limens for vowel formant frequency,'' J. Acoust. Soc. Am. **95,** 1074–1084.

Imai, S., and Abe, Y. (**1979**). ''Spectral envelope extraction by improved cepstrum,'' IEICE, J62-A, 4, 217–228 (in Japanese).

Kewley-Port, D., and Watson, C. S. (**1994**). ''Threshold for formant-frequency discrimination for isolated English vowels,'' J. Acoust. Soc. Am. **95,** 485–496.

Lin, Q. (**1990**). ''Speech production theory and articulatory speech synthesis,'' Ph.D. thesis of KTH.

Markel, J. D., and Gray, A. H. (**1976**). *Linear Prediction of Speech* (Springer-Verlag, New York).

Mermelstein, P. (**1967**). ''On the piriform recesses and their acoustic effects,'' Fol. Phoniat. **18**, 153–167.

Moore, C. A. (**1992**). ''The correspondence of vocal tract resonance with volumes obtained from magnetic resonance images,'' J. Speech Hear. Res. **35**, 1009–1023.

Narayanan, S., Alwan, A., and Haker, K. (**1995**). ''An articulatory study of fricative consonants using magnetic resonance images,'' J. Acoust. Soc. Am. **98,** 1325–1347.

Schroeder, M. R. (**1967**). ''Determination of the geometry of the human vocal tract by acoustic measurements,'' J. Acoust. Soc. Am. **41,** 1002–1020.

Sundberg, J. (**1974**). ''Articulatory interpretation of the singing formants,'' J. Acoust. Soc. Am. **55,** 838–844.

Yang, C., Kasuya, H., Kanou, S., and Satou, S. (**1994**). ''An accurate method to measure the shape and length of the vocal tract for the five Japanese vowels by MRI,'' Jpn. J. Logopedics Phoniatrics **35** 317–321 (in Japanese).