TR - H - 189

Acoustic characteristics of the piriform fossa in models and humans

Jianwu Dang Kiyoshi Honda

1996. 3. 15

ATR人間情報通信研究所

〒619-02 京都府相楽郡精華町光台2-2 ☎ 0774-95-1011

ATR Human Information Processing Research Laboratories
2-2, Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-02 Japan
Telephone: +81-774-95-1011
Facsimile: +81-774-95-1008

ACOUSTIC CHARACTERISTICS OF THE PIRIFORM FOSSA IN MODELS AND HUMANS

Jianwu DANG[1] and Kiyoshi HONDA[1,2]

[1] ATR Human Information Res. Labs., 2-2 Hikaridai Seikacho Soraku-gun Kyoto, 619-02 Japan

[2] University of Wisconsin, 1500 Highland Avenue, Madison, WI 53705-2280 USA

This paper has been submitted for publication

in the Journal of the Acoustical Society of America

ABSTRACT

The piriform fossa forms the bottom of the pharynx and acts as a side branch of the vocal tract. Because of its obscure form and function, the role of the piriform fossa has been underestimated in the traditional models of speech production. This study examines acoustic characteristics of the piriform fossa on speech sounds by means of mechanical modeling, in-vivo experiments and numerical computations. Morphological data of the piriform fossa were obtained from volumetric MRI images for four subjects. Replica models of the vertical part of the vocal tract were developed based on the MRI data from one of the subjects. Acoustic measurements were performed on those models with and without inclusion of the fossa. The results showed that the piriform fossa does not only contribute a local zero to speech spectra, but also affects the lower vowel formants. In-vivo experiments were conducted on human subjects by injecting water into the piriform fossa. The antiresonances were identified as a pattern of increasing trough in speech spectra while decreasing air volume of the cavities. Acoustic simulation based on numerical models demonstrated that the effects of the piriform fossa extend down to the first vowel formants, because the fossa is located near the glottal end of the vocal tract. The antiresonances obtained from these experiments were confirmed to be consistent with those in natural speech within 5%.

PACS number: 43.70.Bk, 43.70.Aj, 43.72.Ct

INTRODUCTION

The piriform fossa is referred to as a small pocket in the pharynx located just above the entrance of the esophagus, forming two bilateral cavities located at the posterolateral aspect of the laryngeal tube. Therefore, this space belongs to the vocal tract. Except for a few of previous studies (e.g., Fant 1960; Sundberg 1974), however, this structure has not been accepted as a functional part of the vocal tract. Why has the acoustic role of the piriform fossa been ignored in the traditional vocal tract model? There seems to be at least two reasons: one is that available morphological data were insufficient to accurately describe acoustic effects of the piriform fossa; the other is that the small side branch was thought to be of minor importance from the perspective of a formal model of the vocal tract.

In earlier studies, Fant (1960) noted an interesting observation on the role of piriform fossa with respect to vowel formants. On the basis of x-ray data, he evaluated acoustic effects of the piriform fossa on the transfer function using a three-parameter vocal tract model in which the fossa was treated as a side branch. The results indicated that formant frequencies were decreased by about 5% for closed vowels and more than 10% for open vowels when the piriform fossa was taken into account. Flach and Schwickardie (1966) performed an experiment to examine the acoustic effect of the fossa by filling the fossa of living subjects with cotton. They concluded that the fossa showed some effects of sound attenuation in a frequency region above 1500 Hz. However, Mermelstein (1967) pointed out that the use of cotton would not eliminate the acoustic effects because of the physical nature of the material, and further stated that the piriform fossa theoretically did not impose any major effects below 4 kHz. Sundberg (1974) estimated the acoustic effects using a mechanical model, which treated the piriform fossa as a uniform tube with the length derived from x-ray data. He demonstrated that the piriform fossa played a significant role in shaping the singing formant between 2 to 3 kHz. The results did not support Mermelstein's conclusion.

Lin (1990) used a numerical model to show that the piriform fossa with a fixed volume and coupling area could reduce formant frequencies to a variable extent depending on the place of vowel articulation. The effects were more prominent for open vowels because of larger reductions in F1 and F2 frequency values.

With the advent of the MRI technique, it is possible that the piriform fossa is exposed in front of our eyes as a three-dimensional image. A number of MRI-based studies have investigated 3D 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 precise measurements of the cross-sectional shape of the vocal tract, and examined the accuracy by matching the MRI-derived transfer functions with real speech spectra for the same subjects. Baer et al. (1991) analyzed area data of the vocal tract with and without the piriform fossa, and compared synthesized vowels with the natural utterances. Their study also demonstrated a reduction in formant frequencies in accordance with Fant's work, even though in their study the volume of the piriform fossa was treated not as a side branch but as a part of the vocal tract. Davies et al. (1993) used the morphological data obtained by Baer et al. to their computation model of the vocal tract, and computed the transfer function in the conditions both with and without the fossa. The F1 and F2 frequencies for the vowel /a/ decreased by about 5% when the fossa was included in the vocal tract as a side branch. All these studies suggest that the piriform fossa does cause significant effects on speech spectra, and that estimation of vocal tract transfer function could be erroneous when the fossa was neglected. In evidence, many studies have shown that MRI-derived transfer function tends to exhibit slightly higher formant frequencies than those from the real speech. Yang et al. (1994) attempted to explain the discrepancy between estimated and measured formants, noting a factor of underestimation of vocal tract length in the MRI data. Judging from the earlier studies, however, the piriform fossa might be a more critical factor to account for the discrepancy.

The previous studies above have shown both agreements and disagreements with respect to the effect of the piriform fossa, and the exact role of this structure is not yet clear in natural speech. It is therefore suggested that the acoustic effects of the piriform fossa need to be examined on the basis of the detailed morphological data. The present study aims to explore the acoustical characteristics of the piriform fossa by means of morphological and acoustic measurements based on MRI data and in-vivo experiments on the human piriform fossa. Volumetric MRI data were obtained for four subjects. Mechanical models were constructed to replicate a partial vocal tract from the glottis to the velum based on the volumetric data. Those models were used to clarify the acoustical effects of the piriform fossa on output sounds by manipulating the air volume of the cavities. A similar approach was repeated on human subjects by injecting water into the fossa during sustaining vowels. Antiresonances derived from the geometric data were compared with those from natural speech for the subjects.

I. MORPHOLOGICAL MEASUREMENT OF THE PIRIFORM FOSSA

As described in the previous section, one of the problems in the earlier studies of the piriform fossa is the lack of detailed morphological data. For this reason, we measured the dimensions of the piriform fossa from MRI data, and developed mechanical and numerical models based on the results of volumetric analysis.

A. MRI measurement of the piriform fossa

Volumetric MRI data were obtained in both the transverse and coronal orientations during sustaining vowels. The standard spin echo method was employed in the scanning. The relaxation time (TR) was 800 ms and the excitation time (TE) was 18 ms. A 25 cm × 25 cm field of view was digitally represented by a 256 × 256 pixel matrix for each slice (see Dang et al., 1994). The volumetric data consisted of 26 slices for the transverse plane and 24 slices for the coronal plane. The thickness of the slices was 0.5 cm for both orientations with neither gap nor overlap. Each slice of the MRI

data was resampled into an image of 250 × 250 pixels, so that each pixel represents a 0.1 × 0.1 cm square. The resampled slices were further constructed into volumetric data of 0.1 × 0.1 × 0.1 cm voxels by means of interslice image interpolation. The procedure was performed by a commercial software (VoxelView) on a workstation (IRIS Indigo). Figure 1 shows a three-plane image of the volumetric MRI data of the vocal tract. The shape and location of the piriform fossa are demonstrated in relation to the entire vocal tract. The slices in the coronal and transverse plans show that the piriform fossa consists of two bilateral cavities located on both sides of the laryngeal tube. The transverse section indicates that the opening of each cavity of the fossa is larger than that of the laryngeal tube.

Volumetric MRI data were obtained from four subjects: three males and one female. Dimensions of the left and right cavities of the fossa were measured for the Japanese vowels /a/, /i/ and /u/ on the basis of the transverse data with reference to the coronal data. Area functions of the piriform fossa are shown in Fig. 2 for the left and right cavities for KH. The areas were measured from the bottom of the cavity to the horizontal boundary between the larynx and pharynx, which is indicated by line aa' shown in Fig. 1. Depth and volume of the piriform fossa for the four subjects are illustrated in Fig. 3. The average depth values over the vowels are about the same across the subjects. The typical depth is about 1.8 cm. HH shows a relatively greater variation in depth across the vowels than the other subjects. The volume ranges from 1.0 to 1.4 cm^3 for the cavities. Individual variation of the volume is also seen among the subjects. Within a subject, the variation of the volume across vowels is the smallest for RY and the largest for SM. The female subject shows a smaller volume than the males. The morphological data do not generally show any significant asymmetry between the left and right sides. The details are shown in Table I.

B. Three-dimensional shape of the piriform fossa

The vertical part of the vocal tract was replicated by a mechanical model to visualize the three-dimensional form of the piriform fossa. The transverse MRI data were represented by layers of vinyl chloride plates with a visual reference to the coronal images. The plates of a size of 7 cm × 7 cm × 0.1 cm were used to replicate the air-tissue pattern of image slices. Three models were constructed from the MRI data for sustained Japanese vowels /a/, /i/ and /u/ that were obtained from KH. The models were about 8.5 cm long, from the subglottis of about 1 cm below the vocal folds to the bend near the velum. Openings of the piriform fossa in the models are at about 3.5 cm from the input end and about 5 cm from the output end. The vallecula is about 6 cm from the input end for the models of vowels /i/ and /u/, while it did not appear in the vowel /a/.

According to the mechanical model, a schematic view of the vocal tract in the vicinity of the piriform fossa is depicted in Fig. 4. Anatomical relations are noted here briefly. The bottom of vocal tract divides into three small branches, the laryngeal tube (vestibule of the larynx) and bilateral cavities of the piriform fossa. The laryngeal tube is a short conduit of about 2 cm long in male, bounded by the vocal folds at the bottom and by aryepiglottic fold at the top. In quiet respiration and phonation, the laryngeal tube opens dorsocranially into the center of the pharyngeal cavity. The piriform fossa is located behind the larynx tube, and forms the bottom of the pharynx just above the esophageal entrance. From a functional point of view, it is a passage of food bifurcating into two small pouches with a funnel shape, one left and one right. A small food bolus reaches here before it is swallowed into the esophagus. The horizontal section shown in the figure indicates that relative relationship of each cavity of the piriform fossa and the laryngeal tube. The cavities open into the main pharynx near the top of the laryngeal tube, sharing the aryepiglottic folds as a common boundary. However, the opening of the piriform fossa is not demarcated by a clear border because the medial wall of the piriform fossa continues anteriorly to the vallecula, which is the space between the tongue root and the epiglottis. With reference to the direction of

sound propagation, the piriform fossa configures a side branch in the vocal tract. Since the shape of the cavities is relatively unvarying during quiet respiration and phonation, the piriform fossa is expected to contribute relatively invariant acoustic features to the transmission characteristics of the vocal tract.

The mechanical models show that the lateral glosso-epiglottic fold and the lateral wall of the pharynx form a half-cylinder-like portion, shaping the opening of the fossa. This portion effectively extends the fossa cavity. A coronal MRI slice and a schematic diagram of the piriform fossa are illustrated in Fig. 5 to show the effective cavity. In this figure, the piriform fossa is shown to consist of two portions: one is the lower portion from the bottom of the fossa to the horizontal boundary of the larynx and the pharynx; the other is the portion superior to the horizontal boundary. Point A is the intersection of the horizontal boundary and the effective radiating end, and Point B is the intersection of the lateral wall and the effective radiating end. L_B represents the length from the boundary to point B. According to the MRI data, L_B can be represented by the following equation,

$$L_B = 1.3/D \qquad (1)$$

where D denotes the diameter of the cross-sectional area of the piriform fossa in the horizontal boundary, and the area is assumed to be circular.

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 acoustic consequence of the geometry, such as the end correction for the cavity opening of the piriform fossa. However, estimation of accurate parameter values is difficult because of the complex shape of the structure. For this reason, mechanical model of the vocal tract was first used in this study. The coefficients for end correction are derived from the experiments on the mechanical model, and used for

numerical simulation in the later sections. The M-model also served in a pilot experiment of water injection in human subjects.

A. Experimental procedure

Treating the mechanical model as an acoustic tube, the input end is the "subglottal" side, and the output end is the "velum" side (see Fig. 6). Since the models represent only a part of the vocal tract, they do not demonstrate resonance properties of the entire vocal tract. As long as the side branches in the models are concerned, however, they exhibit the same antiresonance on the transfer function as they do in the entire vocal tract. Since numerical models are also employed in the later part of this study, a confusion may occur in distinguishing two types of models. For this reason, the mechanical model is referred to as the M-model and the numerical model as the N-model. The M-models are labeled M-model /a/, M-model /i/, and M-model /u/, corresponding to their original vocal tracts, respectively.

The acoustic effects of the piriform fossa in the M-models were examined using a two-point sound pressure method. One of the pressures was inside the model tract near the input end and the other at the radiating end. The two-point pressure method proves that the obtained transfer function precisely represents the acoustic properties of a segment from the inside measurement point to the radiating end, because the effects of the sound source and the geometric shape below the measurement point are excluded from the measurement (Dang & Honda, 1996a). Fig. 6 shows the experimental setup for measuring the transfer function of the M-models. Sound pressures inside and outside the M-model were recorded by two microphones. Microphone M1, a B&K 4003 (1/2 inch condenser microphone), was placed about 6 cm away from the radiating end. Probe microphone M2, a B&K 4182, was used to acquire the intrapressure just above the model's "vocal folds" via a flexible probe tube with a length of 5 cm. The probe tube was installed through a hole opened in the lateral wall of the larynx of the M-model. The probe tube had an outer diameter of 0.165 cm, an inner diameter of 0.076

cm, and a matched impedance to the microphone. A white-noise signal was produced using an FG-143 function generator (NF Circuit Design Block Co.). A horn driver unit, SG-505FRP (Goto Unit Co.), was used to feed the sound into the M-model. The joint between the M-model and the throat of the horn driver was sealed with plasticine to prevent sound leakage.

The M-models were measured under four conditions: both cavities filled with plasticine, either the right or left cavities filled, and both cavities empty. Sound recording was carried out at a sampling rate of 44.1 kHz in an anechoic room. A room temperature of 20 C° was maintained during the measurement, where the sound velocity is expected to be 34300 cm/s. It was confirmed that outside sound pressure in the vicinity of the joint was about 25 dB lower than that at the radiating end. This implies that sound leakage from the joint is negligible in this measurement. An FFT-derived cepstrum analysis (Imai & Abe, 1979) was applied to the recorded signals in order to obtain the envelope of the spectra, where the cepstrum coefficients were weighted by the Hamming window with a length of 0.05 second for white noise and of 1.2 times the fundamental period for speech sound.

B. Changes in the transfer function with the piriform fossa

Pressure-pressure transfer functions of the three M-models were evaluated from the measured sound pressures under the four conditions described above. The results are shown in Fig. 7 for the M-models, (a) with and without the both cavities and (b) with either the left or right cavity alone, respectively.

In Fig. 7(a), the solid lines show the spectra obtained under natural conditions, i.e., with both cavities empty; the dashed 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 the natural conditions for all the M-models of /a/, /i/, and /u/. When the both cavities are filled, the troughs are replaced by peaks at almost the same

frequency region around 4 kHz. The amplitude of spectra in the frequency region is more than 30 dB larger without the piriform fossa than with the fossa. The results indicate that the troughs are definitely caused by the piriform fossa behaving as a side branch. The large changes in the spectra imply that the piriform fossa has a significant effect on the transmission characteristics due to its antiresonance. In Fig. 7(b), the solid lines show the spectra obtained from the M-models with the left cavity only, and the dashed lines indicate the results with the right cavity only. The results show almost identical spectral shapes for the left and right cavities except for a small discrepancy in the trough's shape. This implies that no significant asymmetry exists in the left and right cavities for this subject.

The antiresonances caused by the piriform fossa are summarized in Table II for each condition. The M-model /a/ shows a higher antiresonance frequency than the others. The antiresonance frequency caused by a single cavity was lower than that caused by the both cavities together. The same tendency is seen for all three M-models. This phenomenon might be caused by a stronger end-effect of the opening that occurs for the case of a single cavity, because the area ratio of the fossa opening to the cross-sectional area of the lower pharynx end is larger for a single cavity than for the both cavities together. This speculation suggests a necessity of end correction for incorporating the piriform fossa into the transmission line model. To evaluate radiation coefficient of the piriform fossa, area function of the M-models was used for computing the antiresonances of the fossa, where the fossa was represented by the schematic tube shown in Fig. 5. Length L_B of the superior part of the fossa was derived by Eq. (1). The optimum value was obtained by matching computed antiresonances to those measurements while filling the fossa from 0 to 0.9 cm in a 0.1-cm interval. When changing radiation coefficient from 0.5 to 0.85 by increments of about 0.015, the value of 0.75 met the given condition that computed antiresonances were consistent with those measurements within 4% for the three M-models (Dang at el., 1996b).

Besides the trough near 4 kHz, relatively small but non-negligible effects on the peaks can also be seen in the lower frequency region below 3 kHz. Frequencies of the peaks are listed in Table III. Note that the peaks do represent the acoustic properties of the M-models, but do not correspond to vowel formants, since the M-models represent a partial vocal tract. However, the changes in the peaks approximately demonstrate the size of the effect of the piriform fossa on peaks that are located in the frequency region. As a general tendency in our data, the piriform fossa lowers the frequencies of the peaks that are in the region below the antiresonance frequency of the fossa. For the first peak, the changes in the frequency are 82 Hz for /a/, 94 Hz for /i/, and 94 Hz for /u/. For the second peak, the deviation of the frequency is 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 peaks about 8% in the region around 1 kHz and about 13% in the region above 3 kHz.

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

The direct approach to explore the acoustic effects of the piriform fossa used above is not applicable to living human subjects. However, the similar experiment could be performed by injecting water to the fossa of human subjects. The prediction is that if water is injected in the piriform fossa of human subjects during speech, the effects of the piriform fossa could be observed as increasing antiresonance frequency with decreasing the air volume of the cavity. The feasibility of this idea was tested by a pilot experiment on the mechanical models.

Fig. 8 shows the experimental setup for the pilot experiment on an M-model. A B&K 4003 microphone was 6 cm apart from the radiating end to recorded the output sound. A membrane of polyvinylidene chloride was placed between the M-model and the neck of the driver unit to prevent water leakage into the driver unit. The joint between the M-model and the driver unit was sealed with plasticine. A spongy material was obstructed in the glottal end of the model to reduce effects of changes in air volume

of the piriform fossa by increasing the input impedance of the M-model. The excitation sound source was the same as used above, and the amplitude was fixed during the measurement. Radiated sound of the model was recorded while water was injected into the piriform fossa at a uniform speed via a flexible tube with a 0.3-cm outer diameter.

The spectra of the recorded sound are shown in Fig. 9 for M-model /a/. Each curve of the spectra corresponds to a frame of 46-ms, and the frame-to-frame interval was about 130 ms. Note that the spectra included the source characteristics and vibration of the membrane. Since the properties of both the sound source and the membrane are considered to be constant during the measurement, the variations in the spectra are judged to be caused only by the changes in the air volume of the fossa.

In the initial condition with no water injection, the spectral curves show a constant trough at about 4300 Hz. After about 1.2 seconds, water is injected into the right cavity of the piriform fossa. The trough moves towards higher frequency as the air volume of the right cavity decreases due to water injection. 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 fully filled with water after about 4 seconds. Then, the trough by the left cavity begins to increase as the water flows into the left cavity. Unlike the gradual trough motion for the right cavity, the one caused by the left moves rapidly at the beginning. The cause of this rapid rise of the trough was confirmed by 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. As an antiresonance increases with water injection on one side, the remaining antiresonance shows a slightly lower frequency than that before the water injection. This tendency is the same as that seen in the previous experiment. The first peak increases about 9% when the piriform fossa was completely filled with water.

III. ANTIRESONANCES OF THE PIRIFORM FOSSA IN SPEECH SPECTRA

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

A. Water injection experiment on human subjects

The same procedure of water injection was applied on two male subjects JD and KH with the experimental setup shown in Fig. 10. A flexible tube with a 0.3-cm outer diameter and a 0.2-cm inner diameter was inserted along the nasal floor and passed through the nasopharyngeal port to the piriform fossa. The tip of the flexible tube was placed above the right cavity under CRT monitoring with a flexible endoscope. Water of the body temperature was injected into the right cavity, and it flowed into the left side after the right side became full. Radiated oral sounds of sustained vowels /a/, /i/ and /u/ were recorded during water injection.

The spectra of the vowel /a/ are shown in Fig. 11 for JD. The motion patterns of the troughs shown in this figure are similar to those seen in the M-model. For this subject, a trough appears at about 5200 Hz before the water was injected. After about 0.2 seconds, the trough caused by the right cavity rises as the cavity is filled with water. The trough disappears from the spectra when the right cavity is full, which occurs at about 1 second. The trough by the left cavity begins to rise as the water flows into the left side, which occurs approximately after 1 second, and disappears when both the right and left cavities become full after about 2 seconds. In speech spectra, it is also found that antiresonance of a single cavity is lower than that of the both cavities together.

Unlike the M-model, the first formant of speech sound did not show obvious changes. The stability of F1 may be explained by articulatory adjustment via the

auditory feedback. It is reasonable to speculate that the subjects can adjust articulator positions to compensate for the effects of decreasing air volume of the piriform fossa. Nevertheless, some changes in the first formant can be seen from about 1.0 to 1.2 seconds. It may be that, in this period, the subjects could not compensate completely for the rapid change in the air volume of the left cavity, where the water flow rapidly as the surface tension on the right side was released. The running spectra indicate that the piriform fossa causes two antiresonances around 5 kHz in speech spectra for this subject. Motion of antiresonances was also confirmed in speech spectra for KH using the same procedure.

B. Antiresonance of the piriform fossa in natural speech

As demonstrated above, the piriform fossa causes antiresonances in the spectra of the sustained vowels in 4-5 kHz frequency region. Can such antiresonances also be observed in natural speech? To answer the question, natural vowel utterances were analyzed for four subjects, who were the same ones served in the MRI experiment. Table IV shows the speech materials, which consists of single vowels /a/, /i/ and /u/, and three-vowel utterances. Sound recording was carried out in an anechoic room at a sampling rate of 44.1 kHz, where subjects repeated the speech material twice at a natural speech rate in two body postures: upright and supine. Fig. 12 shows spectra of /iai/ recorded in a supine position for KH. The formants shown in this figure significantly change from /i/ to /a/ to /i/. In contrast, a trough constantly appears at about 4300 Hz for the vowels. With reference to the results obtained from the experiments of water injection and plasticine filling, it is obvious that the trough is the antiresonance caused by the piriform fossa.

The acoustic analysis was performed to examine antiresonances from speech samples that were recorded under the above conditions. Since the supine position was employed in the MRI experiment, the antiresonances were extracted for speech sounds recorded in the same position. In this analysis, running spectra as shown in Fig. 12

were computed for each utterance for all the subject. The values of antiresonance frequency were extracted from a stable segment of the running spectra for each vowel. Fig. 13 shows the antiresonance frequencies for the four subjects. The coefficient of variation (CV, the ratio of standard deviation to the mean) for each vowel is about 5% for all the subjects, except that of /a/ for subject RY, which is 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. These results indicate that antiresonances caused by the piriform fossa consistently appear in the spectra of the vowel sequences.

IV. SIMULATIONS USING NUMERICAL MODELS

The experiments using the M-models and the human subjects showed that antiresonances caused by the piriform fossa are consistently observed in transfer functions of the vocal tract. The results strongly suggest that the piriform fossa should be taken into account in a realistic vocal tract model. On the basis of this finding, the piriform fossa was incorporated into a numerical model of the vocal tract. Simulations using the numerical model were conducted to examine global effects on formants as well as local effects of antiresonance. As mentioned earlier, the numerical model is referred to as the N-model to distinguish it from the mechanical model.

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

The N-model is based on a standard transmission model of the vocal tract. In this model, the piriform fossa is treated as a pair of side branches of the vocal tract, and they are represented by an area function with an interval of 0.1 cm. Each cavity of the fossa was shaped by the model design shown in Fig. 5. The radiation coefficient of the cavity opening of the fossa was taken to be 0.75, as obtained in Section II. The radiation impedance of the vocal tract was approximated by a cascade connection of radiation resistance and radiation inductance (Cassé 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.

Accuracy of the N-model was tested by comparing with the pressure-pressure transfer function of the M-models obtained in Section II. It was confirmed that computations from the N-model were consistent within 1% with the measured values of both resonances and antiresonances below 6 kHz. The N-model was then constructed for the four subjects based on their volumetric MRI data, and antiresonances of the piriform fossa were computed using the N-model. The calculated antiresonances were compared for each subject with those obtained from natural speech, as shown in Fig. 14. The antiresonances predicted by the N-models were consistent with those from the speech within 5%, except for /i/ of subject HH. The result shows that the model gives a realistic description of the antiresonances of the piriform fossa.

B. The global effect of the piriform fossa

The experiments on the M-models and human subjects showed that the piriform fossa not only deforms the spectrum around its antiresonance frequency, but also affects the formants in the lower frequency region. This global effect on vowel formants has been demonstrated in earlier studies, both when the fossa was treated as a side branch (Fant, 1960; Davies et al., 1993), and when the fossa volume was included in the vocal tract (Baer et al., 1991). The questions here are why the fossa causes the global effect on the lower formants in both treatments and how the fossa can be more realistically modeled. To answer these questions, simulations were conducted using the N-model for the extreme vowels /a/ and /i/.

The N-model for the simulations consists of the main vocal tract and a side branch of the piriform fossa. The area function for each part was derived from the volumetric MRI data. The branching point of the fossa to the vocal tract was artificially moved from the glottal end to the lip end, while the area functions of both the main tract

and the branch were kept fixed for each vowel. The results are shown in Fig. 15 for /a/ and /i/, where the area functions were obtained from KH. The approximate location of the branching point is shown on the right side of each panel at the intersection of the spectral curves and the right-hand axis. For the top curve in each panel, the transfer functions of /a/ and /i/ have no branch and thus no zeros. When the side branch is included in the N-model, it causes an antiresonance that remains at about 4200 Hz for /a/ and 4000 Hz for /i/.

As shown in this figure, the formant pattern changes significantly when the branching point is moved along the vocal tract, and the changes show different tendencies for the two vowels. When the branching point is moved in the anterior portion from the lips to 11 cm from the glottis, F2 and F3 change significantly for /i/, while no obvious changes are seen for /a/. In contrast, F2 and F3 vary significantly for /a/ as the branching point is moved in the posterior portion, from about 8 cm to the glottis, but no variation shows in the formants for /i/. In the middle portion from about 8 to 11 cm, F3 changes as the branching point moves for both vowels. The results indicate that the side branch location affects the formants of vowels /a/ and /i/ in opposite ways. The main cause of the inverse pattern appears to be the inverse volume ratios between the anterior and posterior portions for the two vowels. F2 and F3 show larger changes when the branch is located at the portion of constriction. In contrast, F1 decreases monotonically for both vowels as the branching point is moved posteriorly along the vocal tract. F1 decreases by approximately 11% for /a/ and 4% for /i/ when the piriform fossa is moved from the lip end to the actual position of 2 cm from the glottal end. Since the location of the branching point is the only variable factor in this simulation, it is obvious that the changes in F1 depend on the location of the branching point, whereas the other formants change mainly with the volume ratio of the branch and the portion including the branch.

The global effects of the fossa on vowel formants can also be simulated by applying perturbation theory to the same morphological data, where the branch was assumed to be a volume perturbation of the main tract. According to perturbation theory (Schroeder, 1967), effects on formant frequencies are approximately proportional to the volume increment and the square of the sound pressure at the location of the volume. A similar relationship can also be derived from the transmission line theory, in which the effect of a branch on vowel formants is proportional to the admittance of the branch and the square of the sound pressure at the branching point. This indicates that effects of the fossa on the formant frequencies depend on the distribution of sound pressure in the vocal tract regardless of whether the fossa is treated as a side branch or as a volume perturbation. Sound pressure in the vocal tract becomes monotonically higher from the lip end to the glottal end for the first formant frequency, while the sound pressure at other formant frequencies alters in a quasi-cyclic manner. This pressure distribution determines the spectral patterns of vowel formants. According to this theory, the piriform fossa has global effects because it is located near the closed end of the vocal tract, where sound pressure is near a maximum for all formant frequencies.

On the other hand, some differences were also seen between the two methods of simulating the fossa. First, the effect on F1 was stronger when the fossa was represented as a side branch than as the volume perturbation when the branching point was at its actual location. Second, perturbation theory predicts that if a perturbation increases the volume at the radiating end all of formant frequencies will be higher than without the perturbation. This prediction was not born out when the fossa was simulated as a side branch. Third, the two simulation methods showed a significant difference in spectral shape around the antiresonance of the piriform fossa, because no trough appeared in the spectra when the fossa was treated as a volume perturbation. Considering the reality observed in speech spectra, the piriform fossa should be modeled as a side branch of the vocal tract.

V. CONCLUSIONS

The piriform fossa is anatomically part of the vocal tract and behaves acoustically as a side branch. The acoustic role of the piriform fossa has been underestimated because of the conjecture that the small cavity causes only a negligible zero at a higher frequency. This study has clarified the acoustic effects of the piriform fossa on the transfer function and speech spectra by means of experiments and simulations.

Morphological data of the piriform fossa were obtained from volumetric MRI images for four subjects. Mechanical models were constructed based on the data from one of the subjects. The models were used in 3D morphological investigation and acoustic measurements. Results from the experiment of the M-model with and without the piriform fossa showed that the fossa causes troughs in the transfer function of the vocal tract in the frequency region of 4 to 5 kHz. The spectral shape is strongly affected in the vicinity of these antiresonances. The effects are also seen at the spectral peaks, even if those located in the lower frequency region. The morphological and acoustic results did not show any significant asymmetries between the left and right cavities of the piriform fossa. Antiresonances caused by each cavity alone of the fossa were lower than those caused by the both cavities together. This study did not provide a convincing explanation for this phenomenon. It remains to be solved by some other methods, such as finite element method.

The effects of the piriform fossa on speech sound were investigated by an in-vivo experiment of injecting water into the fossa in human subjects and using natural vowel utterances. Antiresonances of the piriform fossa constantly appear not only in sustained vowels, but also in the natural vowel utterances. Computed antiresonances of the piriform fossa using the numerical model were consistent with those obtained from natural speech within 5%. The results indicate that the piriform fossa as a part of the vocal tract constantly affects the spectral characteristics of speech.

Numerical simulations were conducted to demonstrate the global effects of the piriform fossa on the formants. The simulations indicated that the piriform fossa has global effects because it is located near the closed end of the vocal tract. The effects of the fossa on vowel formants can be predicted by perturbation theory, where the fossa is assumed to be a volume perturbation of the main tract. Considering the reality of speech spectra, however, the piriform fossa should be modeled as a side branch of the vocal tract.

ACKNOWLEDGMENTS

The authors would like to express their appreciation to Christine Shadle for her helpful discussions and comments. Furthermore, we would like to thank Hiroyuki Hirai and Naoki Kusakawa for their helpful discussions and assistance in the experiments.

REFERENCES

Baer, T., Gore, J., Grocco, L. C., and Nye, P. W. (1991). "Analysis of vocal tract shape and dimensions using magnetic resonance imaging: vowel," J. Acoust. Soc. Am. 90, 799-828.

Cassé, R., Kergomard, J., and Lurton, X. (1984). "Input impedance of brass musical instruments - comparison between experiment and numerical model," J. Acoust. Soc. Ame., 75, 1, 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, 4, 2088-2100.

Dang, J., and Honda, K. (1996a). "A new method for measuring anti-resonances of the vocal tract transmission characteristics: an experiment study of acoustic tubes," J. Acoust. Soc. Jpn (E)., 93-99.

Dang, J., Kawanishi, Y., and Shadle, C. (1996b). "Estimation of correction coefficient and location of the radiating end of the piriform fossa," Proc. of Spring Meeting of Acoust. Soc. Jpn., 287-288. (in Japanese)

Davies, L., McGowan, R., and Shadle, C. (1993). "Practical flow duct acoustics applied to the vocal tract," in Vocal Fold Physiology: Frontier in Basic Science, (ed. Ingo Titze), 93-142.

Fant, G. (1960). *Acoustic theory of speech production*, Mouton, The Hague, (2nd ed., 1970).

Flach, M. and Schwickardie, H. (1966). "Die Recessus Piriformes unter phoniatrischerSicht," Fol. Phoniat. 18, 153-167.

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

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

Mermelstein, P. (1967). "On the piriform recess 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 imaging," J. Acoust. Soc. Am. 98 (3), 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," The Japan Journal of Logopedics and Phoniatrics, 35 317-321. (in Japanese).

Captions of Tables

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

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

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

Table IV Speech material used in this experiment

Captions of figures

Fig. 1 A three-plane image of the vocal tract and the piriform fossa based on volumetric MRI data.

Fig. 2 Area functions from the opening end to the lower end of the piriform fossa for subject KH: (a) the left cavity and (b) the right cavity.

Fig. 3 Depth (a) and volume (b) of the left and right cavities of the piriform fossa measured from volumetric MRI data of the four subjects.

Fig. 4 A diagram of the vocal tract shape in the vicinity of the piriform fossa. The plate shows a section near the opening end of the piriform fossa.

Fig. 5 A coronal slice of the piriform fossa (a) and a diagram of the effective cavity of the piriform fossa (White arrows indicate the horizontal boundary of the pharynx and larynx; L_B is the length of the portion superior to the horizontal boundary).

Fig. 6 The experimental setup to measure the acoustic effects of the piriform fossa using a mechanical model constructed from volumetric MRI data.

Fig. 7 Transfer function, radiated pressure(Pr)/intrapressure(Pi), of the mechanical models when one or both cavities are empty, or filled with plasticine.

Fig. 8 The setup of the pilot experiment to observe the changes in antiresonances when injecting water into the piriform cavities of the M-models.

Fig. 9 Spectra of the radiated sound from the M-model /a/ when excited with a constant sound source and while injecting water into the piriform fossa.

Fig. 10 Experimental setup to inject water into the piriform fossa of human subjects, and thus observe the effects of the piriform fossa on speech sound.

Fig. 11 Spectra of vowel /a/ produced by subject JD while water is injected into the piriform fossa.

Fig. 12 Spectra of /iai/ uttered by subject KH at normal speed and normal pitch in a supine position.

Fig. 13 Antiresonance frequencies caused by the piriform fossa in natural speech of the four subjects.

Fig. 14 Comparison of the antiresonance obtained from the N-models and from natural speech.

Fig. 15 Transfer functions, radiating-pressure/glottal-velocity, of the vocal tract of (a) /a/ and (b) /i/ with and without the piriform fossa, based on the area function derived for subject KH. Uppermost curve in each graph shows spectrum without the piriform fossa. All other curves show spectra with the piriform fossa at different branching points. The lowest curve in each graph is the spectrum when the piriform fossa is in its actual location.

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

Subject	Vowel	Vr(cm^3)	Vl(cm^3)	Dr(cm)	Dl(cm)
HH	/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	/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	/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	/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

Table II Antiresonance frequencies caused by the empty cavities of the pyriform 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

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

Table IV Speech material used in this experiment

Sustained vowels	VVV utterances
/a/, /i/, /u/	/aia/, /aua/, /iai/, /iui/, /uau/, /uiu/

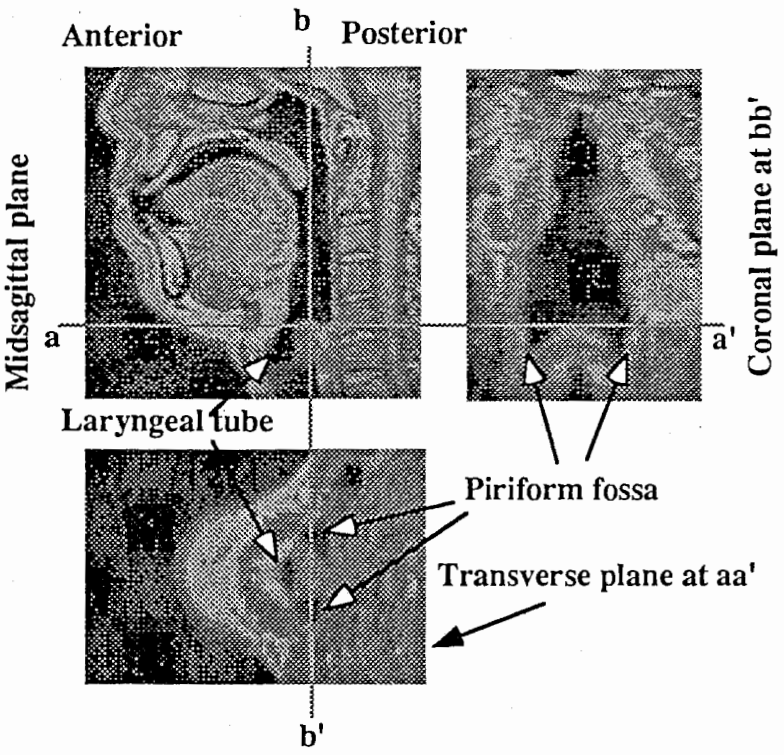


Fig. 1 A three-plane image of the vocal tract and the piriform fossa based on volumetric MRI data.

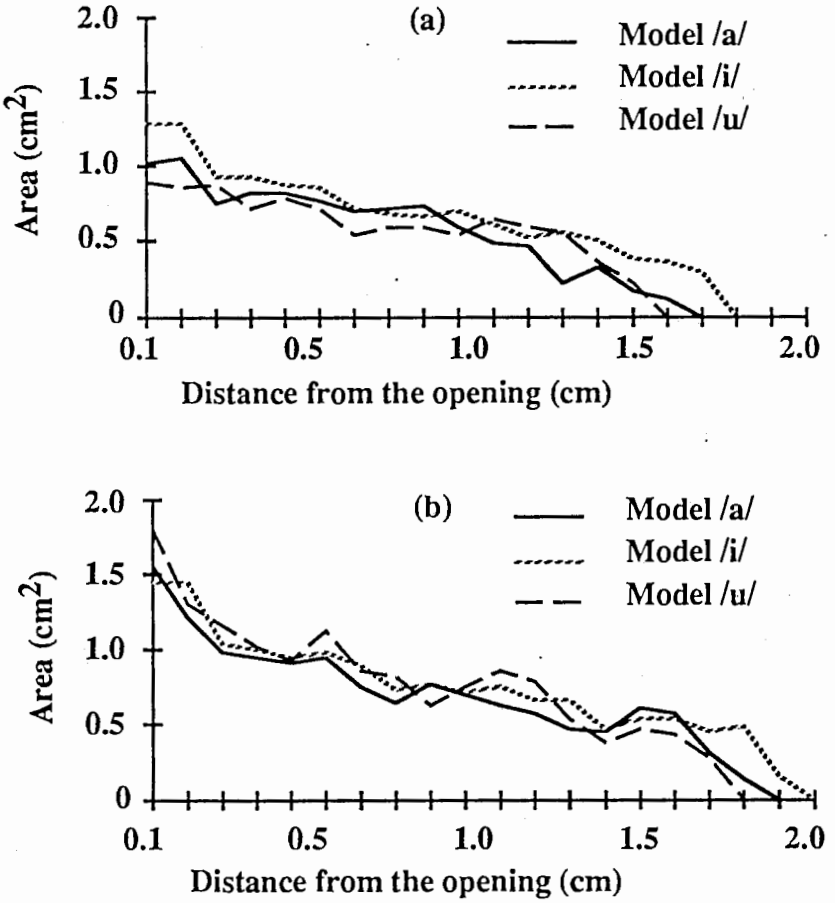


Fig. 2 Area functions from the opening end to the lower end of the piriform fossa for subject KH : (a) the left cavity and (b) the right cavity.

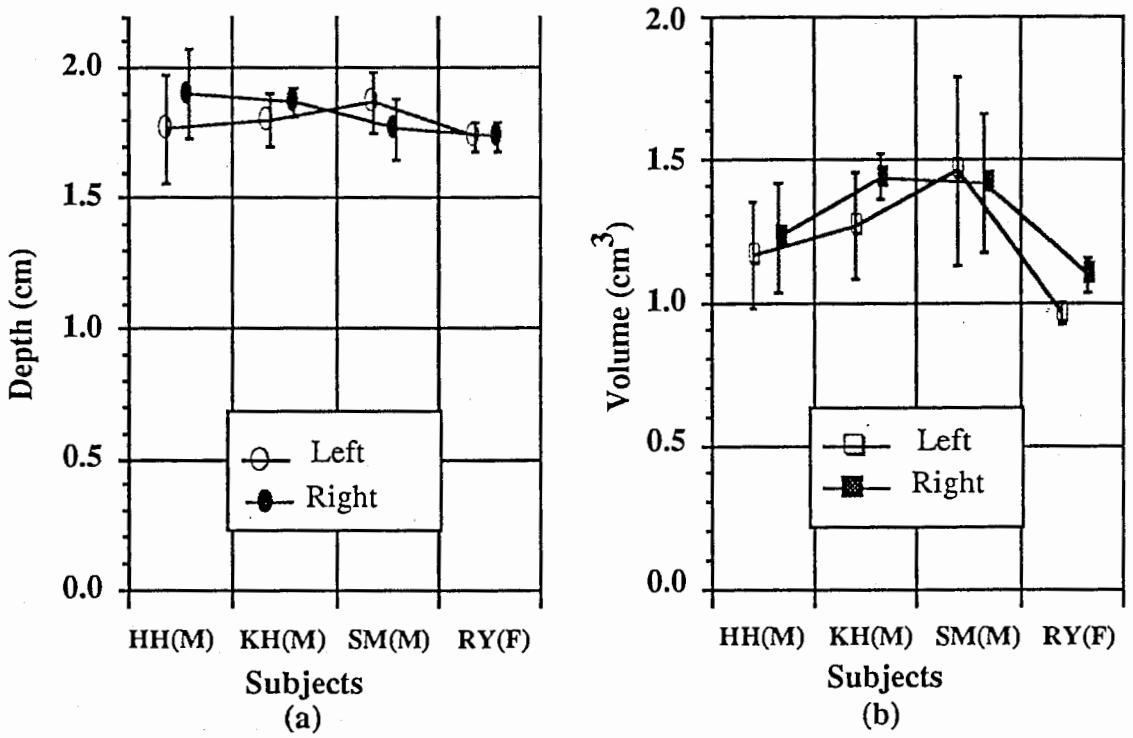


Fig. 3 Depth (a) and volume (b) of the left and right cavities of the piriform fossa measured from volumetric MRI data of the four subjects.

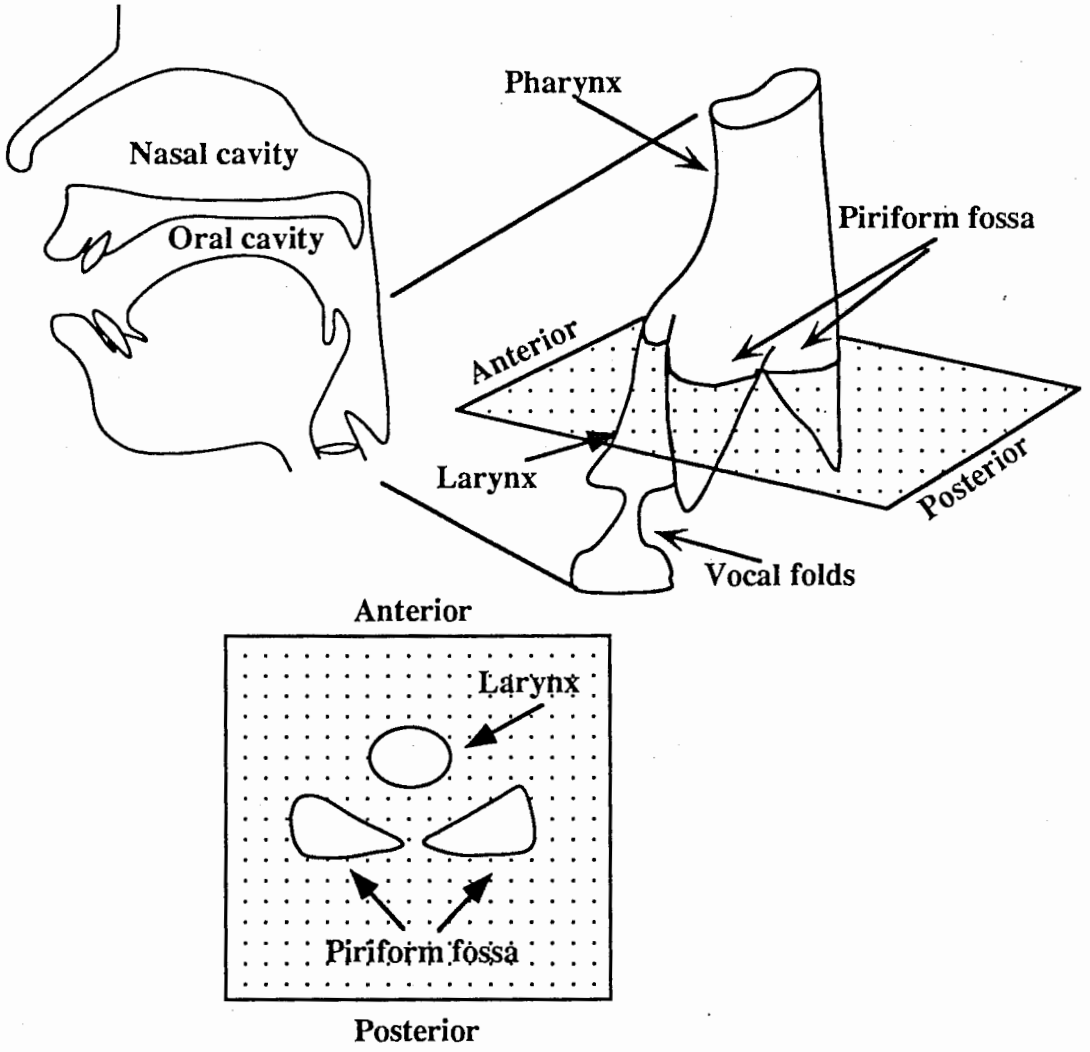
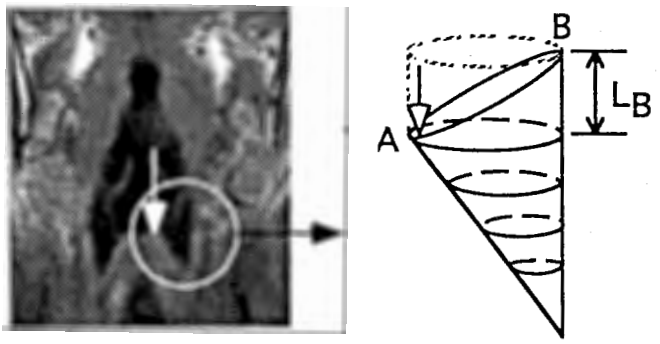


Fig. 4 A diagram of the vocal tract shape in the vicinity of the piriform fossa. The plate shows a section near the opening end of the piriform fossa.



(a) (b)

Fig. 5 A coronal slice of the piriform fossa (a) and a diagram of the effective cavity of the piriform fossa. (White arrows indicate the horizontal boundary of the pharynx and larynx; L_B is the length of the portion superior to the horizontal boundary).

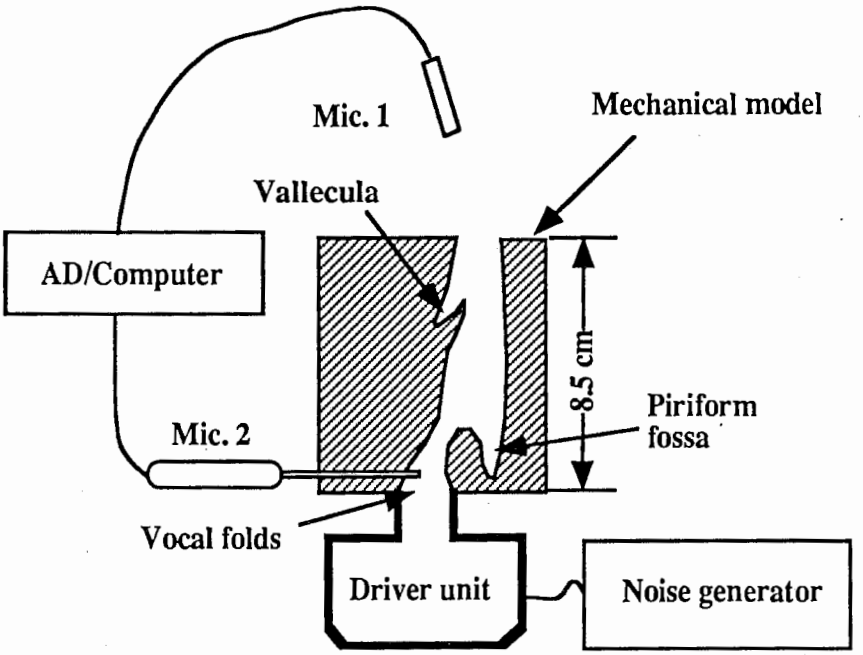


Fig. 6 The experimantal setup to measure the acoustic effects of the piriform fossa using a mechanical model constructed from volumetric MRI data.

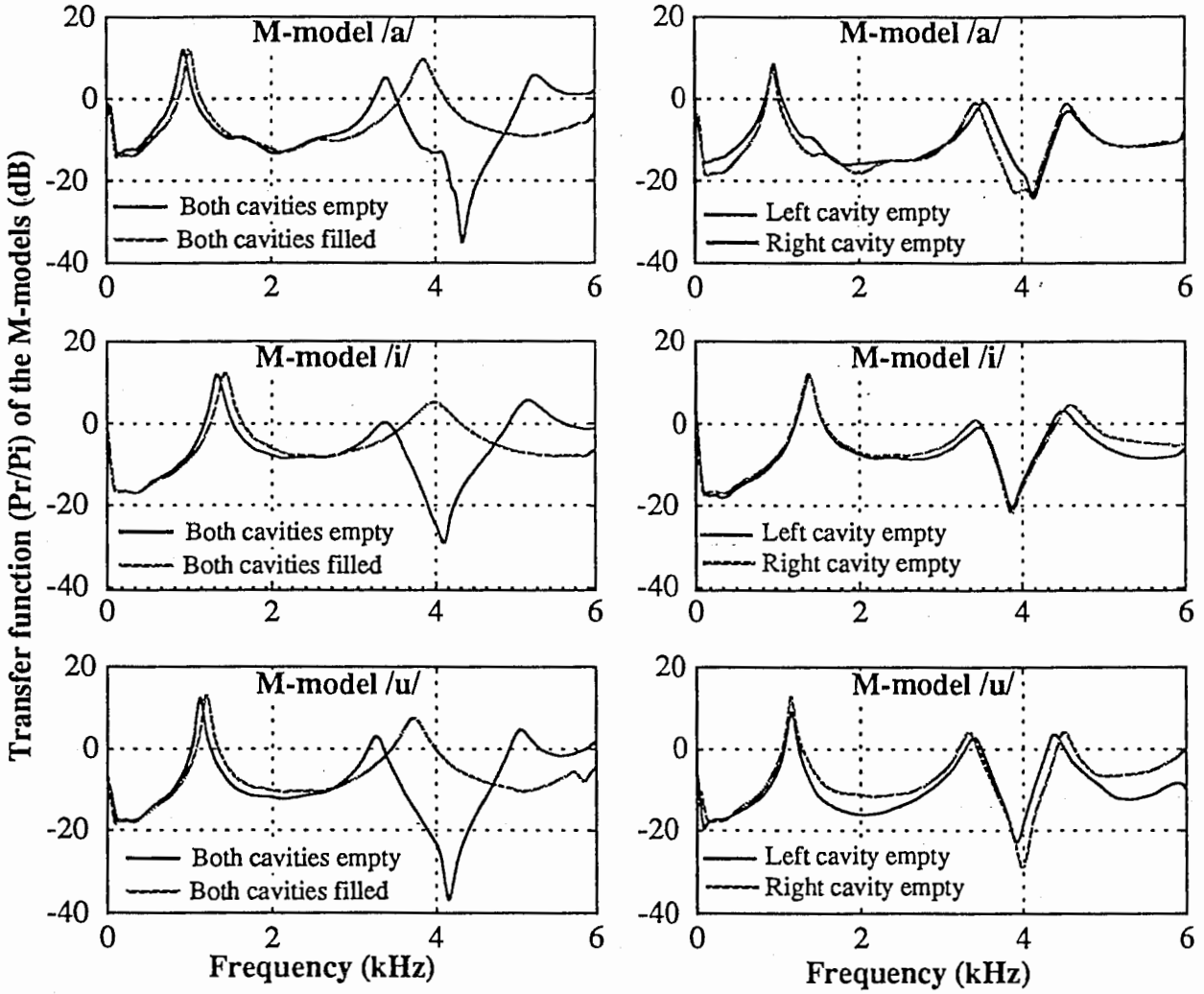


Fig. 7 Transfer function, radiated pressure(Pr)/intrapressure(Pi), of the mechanical models when one or both cavities are empty, or filled with plasticine.

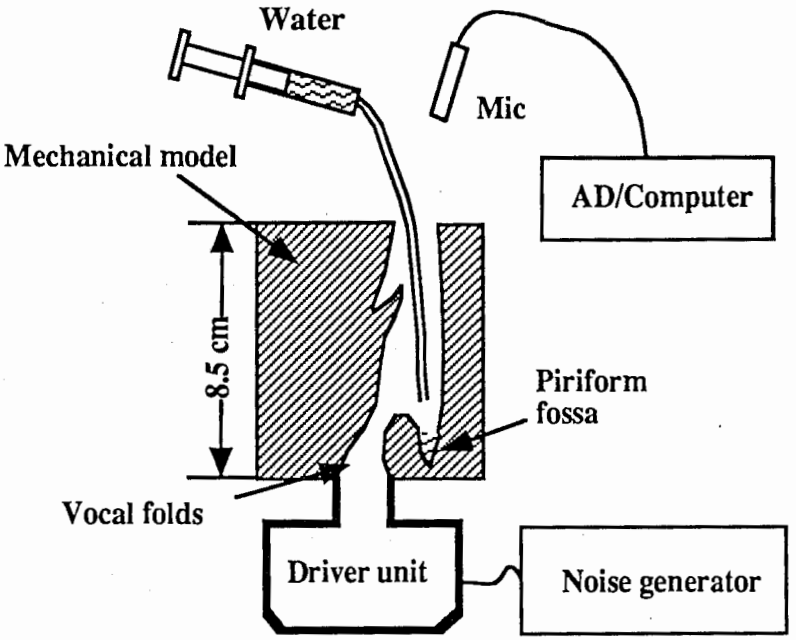


Fig. 8 The setup of the pilot experiment to observe the changes in antiresonances when injecting water into the piriform cavities of the M-models.

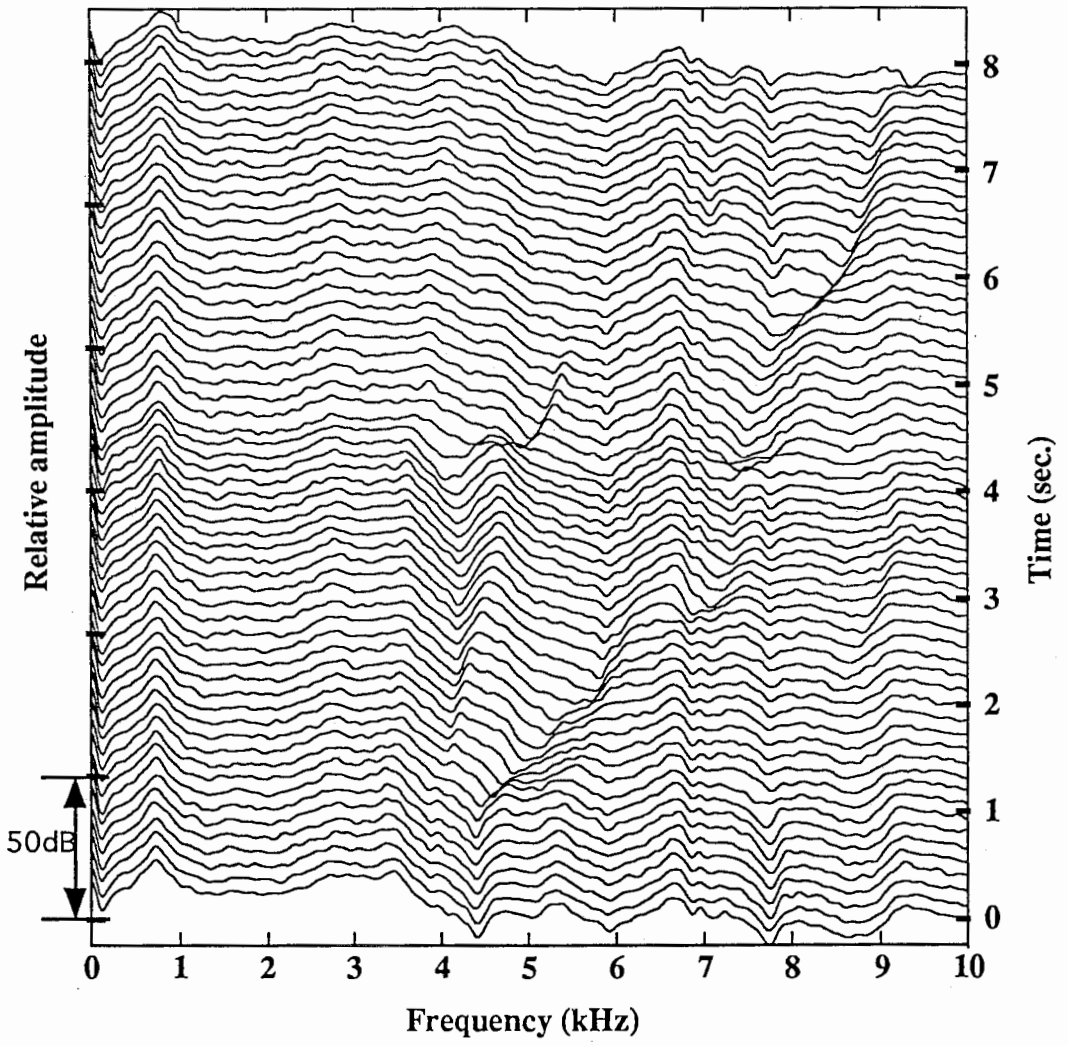


Fig. 9 Spectra of the radiated sound from the M-model /a/ when excited with a constant sound source and while injecting water into the piriform fossa.

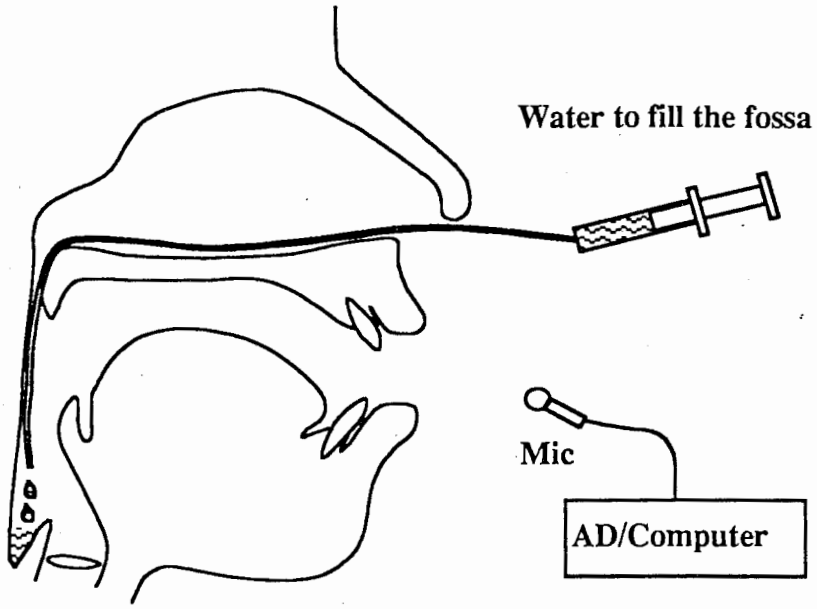


Fig. 10 Experimental setup to inject water into the piriform fossa of human subjects, and thus observe the effects of the piriform fossa on speech sound.

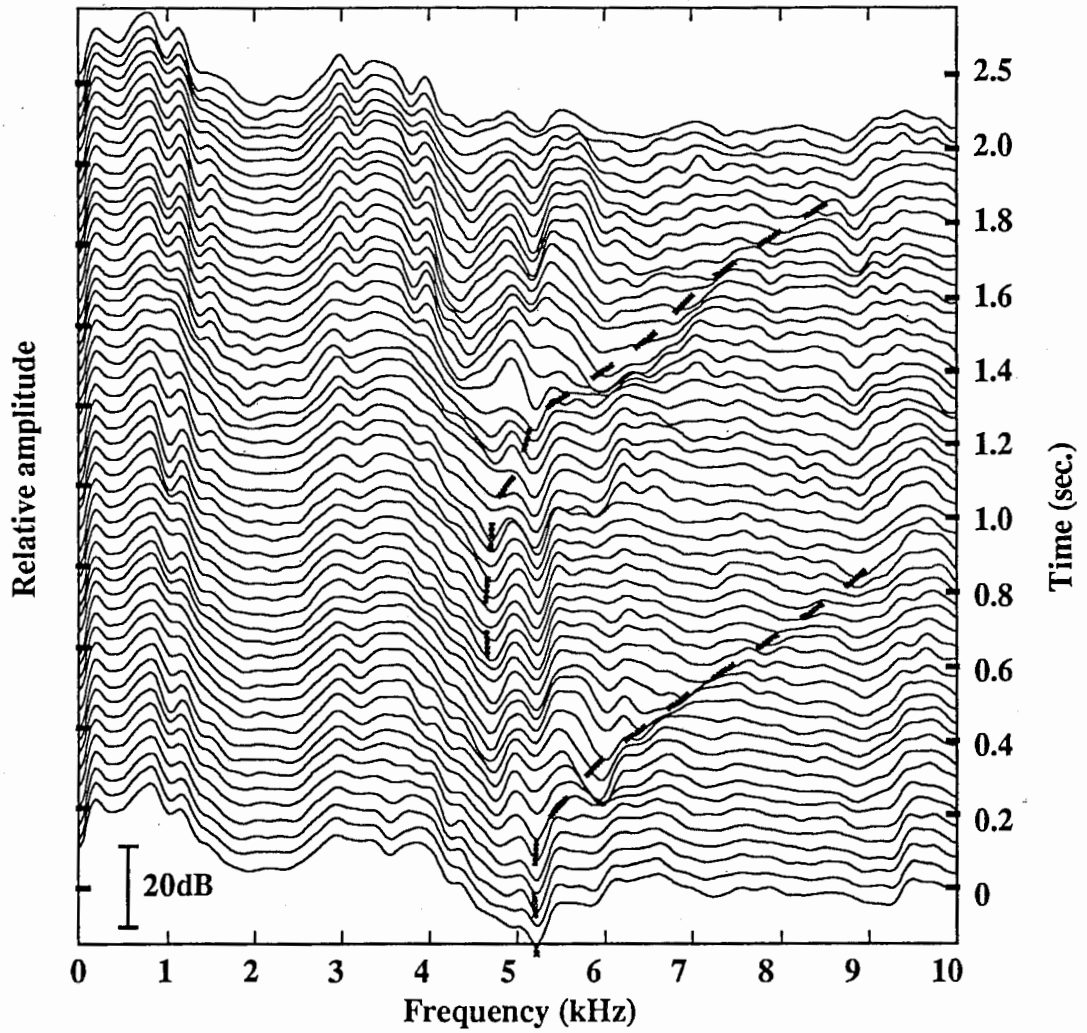


Fig. 11 Spectra of vowel /a/ produced by Subject JD while water is injected into the piriform fossa.

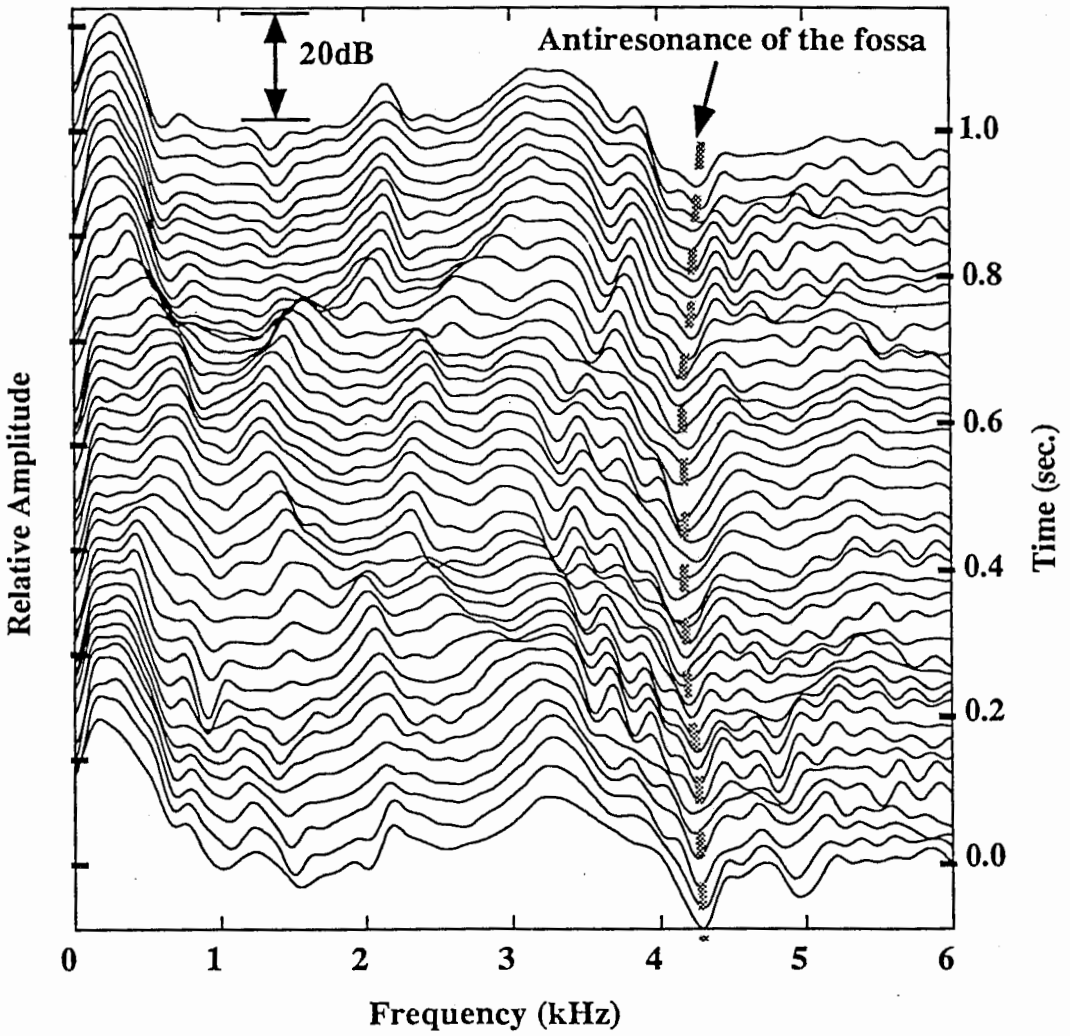


Fig. 12 Spectra of /iai/ uttered by subject KH at normal speed and normal pitch in a supine position.

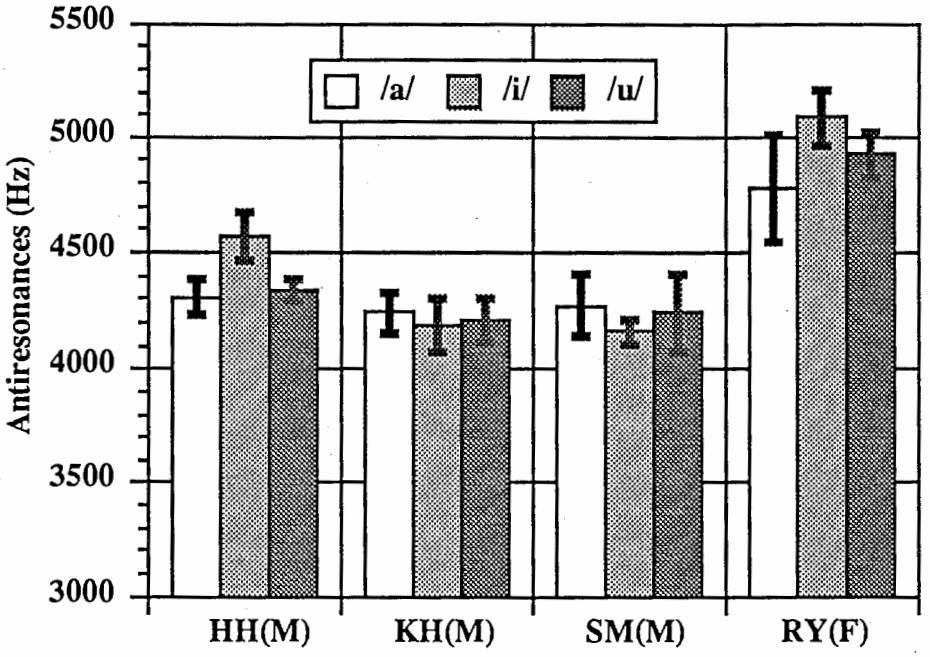


Fig. 13 Antiresonance frequencies caused by the piriform fossa in natural speech of the four subjects.

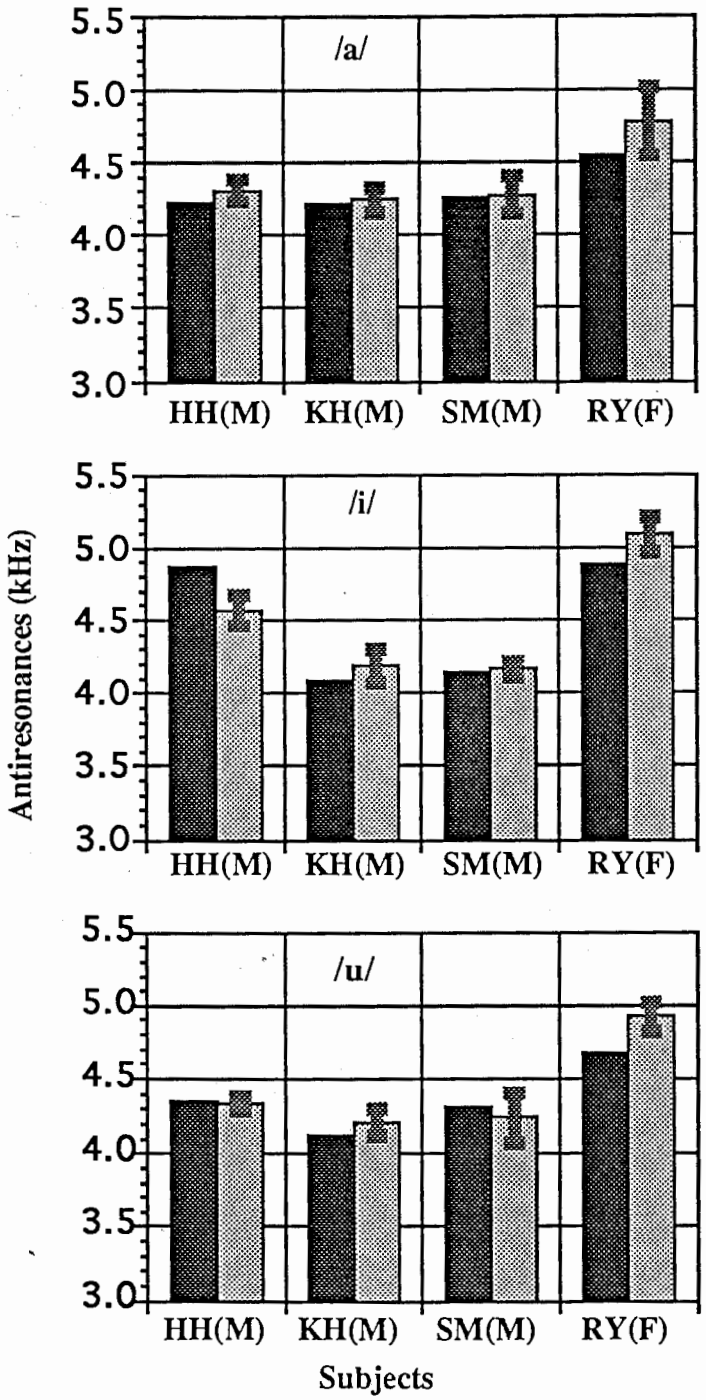


Fig. 14 Comparison of the antiresonance obtained from the N-models and from natural speech.

from natural speech from the N-model

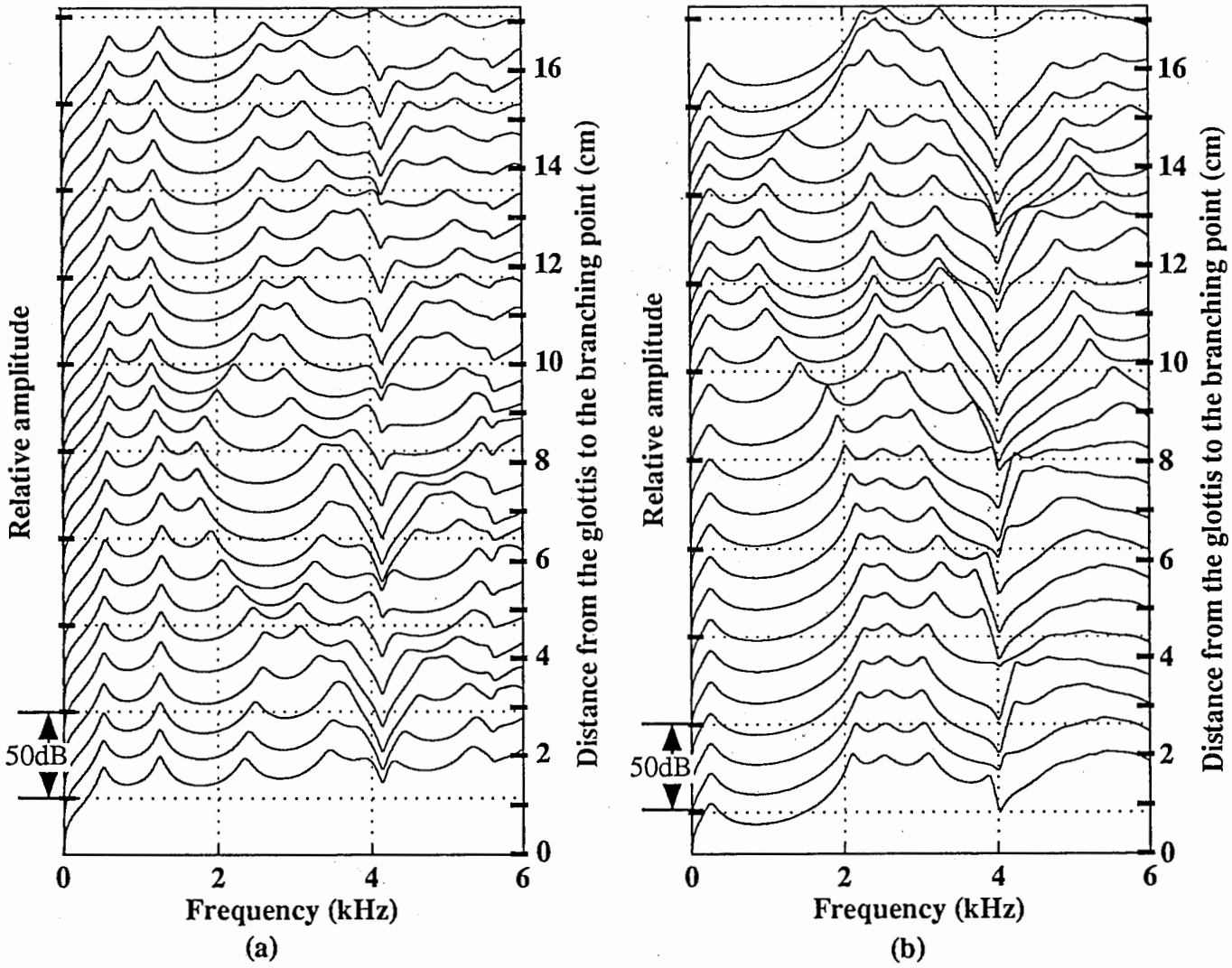


Fig. 15 Transfer functions, radiating-pressure/glottal-velocity, of the vocal tract of (a) /a/ and (b) /i/ with and without the piriform fossa, based on the area function derived for subject KH. Uppermost curve in each graph shows spectrum without the piriform fossa. All other curves show spectra with the piriform fossa at different branching points. The lowest curve in each graph is the spectrum when the piriform fossa is in its actual location.