# AN IMPROVED VOCAL TRACT MODEL OF VOWEL PRODUCTION IMPLEMENTING PIRIFORM RESONANCE AND TRANSVELAR NASAL COUPLING

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

[1]ATR Human Information Processing Res. Labs., 2-2 Hikaridai Seikacho Soraku-gun Kyoto, 619-02 Japan
[2]University of Wisconsin, 1500 Highland Avenue, Madison, WI 53705-2280 USA

## ABSTRACT

This paper proposes an improved vocal tract model of vowel production, which incorporates acoustic effects of the piriform fossa and transvelar nasal coupling. In this study, the vocal tract model was derived from the MRI data of a subject. The piriform fossa was modeled based on the MRI data as a side branch of the vocal tract. The velum wall was modeled as a cascaded impedance of a viscous resistance, a mass and a stiffness, which were estimated by acoustic and mechanical experiments conducted on three subjects. Transfer functions of vowels /a/ and /i/ were computed under the conditions of with and without the piriform fossa and transvelar nasal coupling. The results showed that both the piriform fossa and transvelar coupling play important roles in shaping the first two formants for closed vowels, while the piriform fossa is the main factor affecting the formants of open vowels. By comparing computed transfer functions with real speech spectra for the same subject, it is clarified that our improved model gives a more realistic performance than the traditional model.

## 1. INTRODUCTION

Models of vowel production have successfully treated the vocal tract as a single tube with varying forms, while leaving physical consequences of the detailed morphology obscure. One of them, for example, is the piriform fossa which consists of a pair of cavities near the glottis. It has been shown that the fossa lowers F1 about 10% for open vowels and about 4% for closed vowels, and certain perceptual effects on vowel quality have also been noted [1, 3, 4]. However, the piriform fossa has not been accepted as a functional part of the vocal tract in the traditional vocal tract model.

Another factor that needs to be considered for vowel production is the transvelar coupling between the oral and nasal cavities. While nasal coupling normally occurs via an open velopharyngeal port, it also takes place via velum vibration. Castelli et al. [5] noticed that a peak at 250 Hz often appears in a vowel transfer function, and suggested that the spectral peak in oral vowels may reflect small degree of nasal coupling. One of the authors has examined acoustic effects of nasal coupling on vowel production, by measuring three isolated sound pressures radiated from the lips, nostrils and pharynx wall [6, 7]. The results showed that considerably large sounds radiate from the nostrils during closed vowels and voiced stops, even when the velum is expected to be closed. In the case of /i/, for example, the sound from the nostrils was only 2 dB smaller than that from the lips at the first formant of about 250 Hz. The result implies that the velum yields sound transmission from the oral cavity to nasal cavity by its vibration, and that the nasal coupling is expected to affect the first formant of closed vowels.

Summarizing the previous studies, the coupling of the nasal and oral cavities adds a larger effect on closed vowels, while the piriform fossa has a significant effect on open vowels. Therefore, in this study, we aim to improve the traditional model by taking both factors into account. Two experiments were conducted to acquire acoustic and morphological clues. Comparisons were made between computed transfer functions and real speech spectra for the same subject.

## 2. PROCEDURE OF EXPERIMENTS

Two experiments were conducted in this study: MRI measurement for acquiring morphological data of the vocal tract, and physical measurements for estimating the coupling parameters of the yielding velum.

### 2.1. Measurement of vocal tract morphology

Volumetric MRI data were obtained in both the transverse and coronal orientations during sustained vowels. The standard spin echo method was employed in the scanning. A 25 cm × 25 cm field of view was digitally represented by a 256 × 256 pixel matrix for each slice (ref. [8]). The volumetric data consisted of 30 slices for both the transverse and coronal planes. The slices were 0.4 cm in thickness for both orientations with neither gap nor overlap. Since the teeth were invisible on the MRI image, they could not be distinguished from the vocal tract air space. For this reason, the space occupied by the teeth was extracted from coronal scans of a collapsed vocal tract, where the subject kept the tongue and lips in tight contact with the dental arch. The extracted images of the teeth were mapped onto vocal tract images of vowels.

The area function was obtained by computing cross-sectional areas on a series of planes that were perpendicular to the vocal tract midline. To do this, the tissue-air boundaries of the vocal tract were traced manually on the mid-sagittal slice, and the mid-points of the shortest line segments between two opposing boundaries were connected to form the vocal tract midline. Then, a series of slice lines that were perpendicular to the vocal tract midline were computed at a constant distance along the midline, and new images on the slice planes were reconstructed from the volumetric images. Area measurement was performed on the reconstructed images to obtain the area function. To minimize errors in the area measurement, volumetric images from the transverse and coronal slices were used to compute vertical and horizontal portions of the vocal tract, respectively. The vertical portion from 0 to 7.5 cm and the horizontal portion from 7.5 to the lips were connected to accomplish the final result.

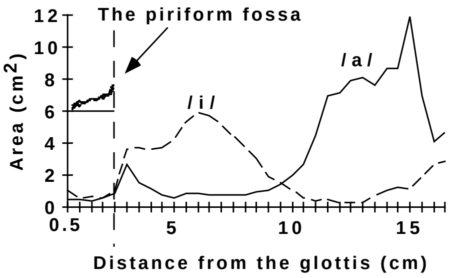

**Figure 1:** Area function of vowels /a/ and /i/ obtained from volumetric MRI data for Subject JD. The left and right cavities of the piriform fossa are plotted together on the middle left side

## 2.2. Estimation of acoustic parameters of the yielding velum

As suggested in previous studies [6, 7], velum vibration is a possible route of oral-nasal coupling in non-nasal vowels. To confirm this proposition, characteristics of velum vibration were measured by an experiment on three subjects. A diagram of the experimental setup is shown in Fig. 2. In this measurement, five channels were used for recording three sound pressure signals and two acceleration signals. M1, a B&K 4003 microphone, was placed in front of the subject (15 cm apart) to record the radiated sound pressures. M2 and M3, two B&K 4182 probe microphones, were used to measure the intraoral and intranasal sound pressures via two identical flexible tubes of 30-cm length. The probe tubes had a 0.165-cm outer diameter, a 0.076-cm inner diameter, and a matching impedance to the microphones. The probe tube of M2 was inserted into the oral cavity and glued onto the hard palate, where its tip was placed beneath the velum. The probe tube of M3 was inserted along the nasal floor through one nostril into the nasopharynx about 7.5 cm back from the nostrils. Mucous clogging of the tube was a potential factor to interfere with the pressure recording. To prevent this, the microphone signals were continuously monitored throughout the experiments so that when mucous clogging occurred it was quickly removed by injecting air into the probe tube. Two accelerometers were used to measure the vibration of the velum (A1) and the lateral surface of the nostrils (A2). The accelerometers were ENDEVCO Model-22 devices with a weight of 0.14 grams. A1 was placed in contact with the velum surface at about the central position on the nasal side by using its wire's stiffness. A2 was attached on the lateral surface of the nostrils to evaluate the sound radiation from the nostrils.

Three male subjects from 30 to 40 years old participated in this experiment. The speech material was Japanese vowels and phrases consisting of the vowels with stop consonants. The analysis was focused on vowel segments.

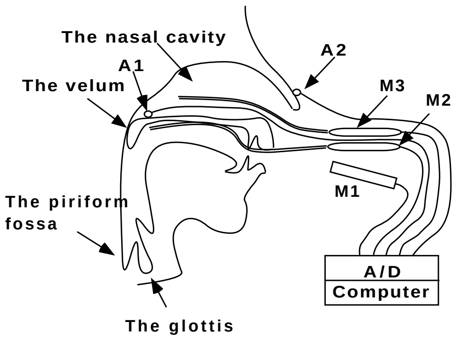

**Figure 2:** A diagram of the setup to measure internal and external sound pressures of the vocal tract and accelerations of the velum and the vocal tract wall.

From this measurement, three physical quantities were acquired for evaluating the velum vibration: volume velocity, input pressure and output pressure. The volume velocity passed through the closed velum was obtained by integrating the acceleration signal of the velum over time. The intraoral sound pressure was considered to be the input to the velum, and the intranasal pressure was the output, that is, the product of the volume velocity and the driving impedance of the nasal cavity looking from the velum. Figure 3 shows spectra of the intraoral and intranasal sound pressures and the acceleration of the velum averaged over all vowel segments. In the figure, the input signal shows a gradually declining spectral contour without large undulation. The remaining signals reach the noise level at 3 kHz for the intranasal pressure, and at 1 kHz for the acceleration.

Considering the velum as the cascade impedance of a mass, a viscous resistance, and a stiffness, the three parameters can be estimated from the volume velocity, input pressure and output pressure using our proposed method [9]. In the frequency range of from 200 to 600 Hz, relatively consistent values were obtained for the three subjects. The equivalent acoustic impedance is

$$Z_v = 80 + jw0.03 - j50000/w \qquad (1)$$

## 3. MODELING

In this study, the piriform fossa is modeled as a side branch of the vocal tract. Our previous study [2] had shown that the effective cavity of the piriform fossa can be described by two vertical portions: a lower portion from the bottom of

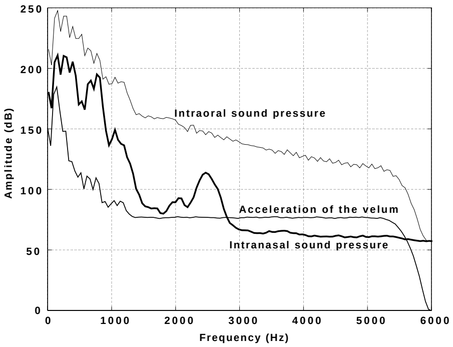

**Figure 3:** The spectra of intraoral, intranasal pressures and the velum acceleration averaged over all vowel segments.

the fossa to the horizontal boundary plane of the larynx and pharynx, and an additional portion above the the horizontal plane. The latter is a half-cylinder-like portion formed by the surrounding structures: the aryepiglottic fold and the lateral wall of the pharynx. The volume of the additional portion can be approximately given by an empirical formula

$$V = 0.65S/D \qquad (2)$$

where S denotes the cross-sectional area of the piriform fossa at the horizontal plane, and D is the diameter where the area is assumed to be circular. In the vocal tract model, the additional portion of the fossa was subtracted from the main vocal tract to avoid the partial volume being doubly used. The open end correction of the effective cavity was 0.75 in this model.

In speech production, the function of the velum is not a binary switch of on and off. In fact, the coupling of the nasal cavity to the oral cavity can be considered to take place via two routes. One is the air passage, when the velum lowers and the velopharyngeal port opens for nasal sounds. The other route is via the velum vibration in non-nasals. The air passage is the main channel between the oral and nasal cavities for nasal sounds, while the velum vibration is the channel for non-nasalized sounds. To include the velum vibration in our vocal tract model, the acoustic impedance of Eq. (1) is connected between the oral cavity and the nasopharynx, shunted with the velopharyngeal port.

# 4. SIMULATION AND ASSESSMENT

The effects of the piriform fossa and the transvelar coupling were simulated using models with and without those factors. Comparisons were made between computed transfer functions and real speech spectra to evaluate the models.

## 4.1. Simulations using the models

The test models used in this computation were derived from the area functions of /a/ and /i/ shown in Fig. 1 with and without the two factors discussed above. The vocal tract wall impedance proposed by Flanagan et al. [10] was employed in the models. For convenience, the model of the oral tract in the traditional account is referred to as the *base model*, and the one including the oral cavity with both the piriform fossa and transvelar coupling is called the *improved model.* In the improved model, the nasal tract was adopted for the area function obtained in [8], and the paranasal sinuses were treated as a four-zero model [2].

Computed transfer functions from the models are shown in Fig. 4 for /a/ and Fig. 5 for /i/. The top curves were obtained from the base model. The second curves were computed using the base model with the transvelar coupling. The transvelar coupling increases the first and third formants of /a/ about 1%, and causes two pole-zero pairs at about 300 Hz and 3000 Hz. When the piriform fossa is taken into account, F1 and F3 of /a/ decrease about 15%, and F2 and F4 about 5%.

For /i/, on the other hand, the transvelar coupling lowers F1 about 6%, while no effect appears in F2. In contrast, the fossa lowers F2 about 16%, and only 1% for F1 and F3. The results indicate that the piriform fossa has significant effects on open vowels, while the transvelar coupling plays an important role in the production of closed vowels.

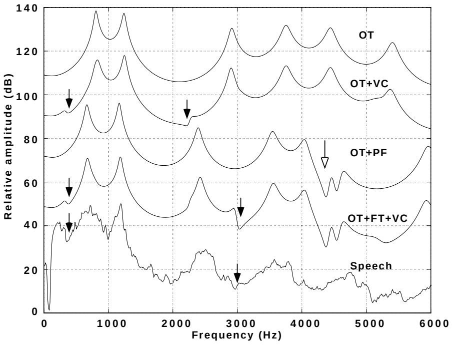

**Figure 4:** Computed transfer functions and spectrum of real speech for /a/. (OT: Oral tract; PF: Piriform fossa; VC: Velum coupling.)

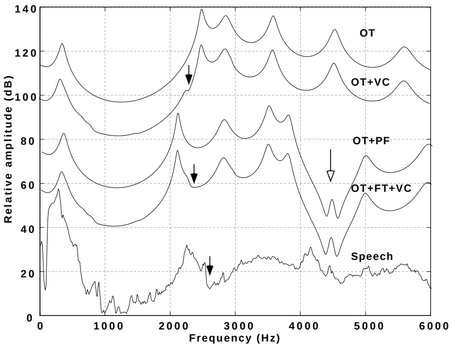

**Figure 5:** Computed transfer functions and spectrum of real speech for /i/. (OT: Oral tract; PF: Piriform fossa; VC: Velum coupling.)

## 4.2. Evaluation of the improved model

Transfer functions computed from the improved model were compared with real speech from the same subject (JD). For

**Table 1:** Variations of the base model and improved model compared to the real speech.(N-model: The base model; I-model:The improved model)

| | F1 | F2 | F3 | F4 |
|---|---|---|---|---|
| Speech /a/ (Hz) | 714 | 1183 | 2472 | 3562 |
| N-model a (%) | +13 | +5 | +17 | +5 |
| I-model a (%) | -4 | 0 | +3 | 0 |
| Speech /i/ (Hz) | 300 | 2250 | ? | ? |
| N-model i (%) | +14 | +10 | - | - |
| I-model i (%) | +8 | -4 | - | - |

an accurate estimation of transmission characteristics of the vocal tract, the subject produced vowels with gradual F0 changes while maintaining the same vocal tract configuration as best as possible. This procedure eliminates the effect of the fundamental frequency on spectral measurement. Each phonation was sustained for about six seconds and repeated for three minutes. The speech sound was recorded at a sampling rate of 48 kHz in an anechoic room. Power spectra were computed on preemphasized signals using a 4096-point FFT with a 1024-point shift, and averaged over the repetitions. The real speech spectra were plotted in the lowest parts of Figs. 4 and 5. Some formant peaks were blurred to some extent perhaps due to the effect of F0 control manuevers on vowel articulation. This can be seen in F3 of the vowels. For vowel /a/, F3 was estimated from an FFT-based cepstrum, so four formants were obtained for /a/. However, only two formants were obtained for /i/. Table I shows the results of the base and improved models compared with the real speech. From Table I, it can be summarized that the improved model is better than the base model in all of the formants, and the improvements are from 5% to 14%.

In Figs. 4 and 5, we use black arrows to show the troughs that were induced by the transvelar coupling, and use white arrows for the troughs caused by the piriform fossa. The troughs are observed consistently in the improved model and the real speech. Castelli et al. showed their measurements of vocal tract transfer functions for vowels [5]. They noted that a peak at 250 Hz often appears in a vowel transfer function, and the 250 Hz peak is clearly seen for open vowels since their first vowel formant is higher. Both our measurements and simulations support their observations. Our simulation showed that pole-zero pairs are obviously caused by the nasal coupling. In addition, the pole-zero pairs are also seen around 2.3 kHz and 3 kHz in both the improved model and the real speech. Such pole-zero pairs can also be seen in literature [5]. In particular, the trough at about 3 kHz constantly appears in the real spectra for other vowels. This implies that the nasal coupling of the velum vibration always acts on oral vowel sounds.

## 5. CONCLUSION

In this study, a side-branch model of the piriform fossa and transvelar coupling of the oral and nasal cavities were combined to formulate a realistic model of the vocal tract. MRI measurement was made to acquire the morphological data of the vocal tract, and acoustic and mechanical measurements were conducted to estimate acoustic parameters of the yielding velum. The results indicated that both the piriform fossa and transvelar coupling play important roles in shaping vowel formants. For closed vowel /i/, the transvelar coupling mainly affects the first formant and the piriform fossa the second one, while the piriform fossa is the main factor for open vowels. In simulation of real speech spectra, the improved model showed a plausible performance compared with the traditional model.

## ACKNOWLEDGMENT

The authors would like to thank N. Kusakawa of ATRI for his help in the acoustic experimental setup, and Y. Shimada and I. Fujimoto of the Takanohara Central Hospital for their assistance in collecting the MRI data.

## REFERENCES

[1] Dang, J. and Honda, K. (1995). "Local and global effects of the piriform fossa," J. Acoust. Soc. Am. 98, p.2931.

[2] Dang, J. and Honda, K. (1996). " Acoustical Modeling of the Vocal Tract based on Morphological Reality: Incorporation of the Paranasal Sinuses and the Piriform Fossa," ETWR-96, (Autrans, France).

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

[4] 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.

[5] Castelli, E., Perrier, P., & Badin, P. (1989). "Acoustic considerations upon the low nasal formant based on nasopharyngeal tract transfer function measurements," Proc. of ECSCT, Vol. 2, 412-415 (Paris).

[6] Suzuki, H., Dang, J., & Nakai, K. (1991). "Measurement of sound vibration at the lips, nostrils and pharynx wall in speech utterance and simulation of sound leakage from the oral cavity to the nasal cavity in non-nasal sound," Jpn. IEICE Trans., J74- A,1705-1714 (in Japanese).

[7] Dang, J., Nakai, K., & Suzuki, H. (1993). "Measurement and simulation of intraoral pressure and radiation of stop consonant," J. Acoust. Soc. Jpn., 49, 313-320 (in Japanese).

[8] 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.

[9] Dang, J., Nakai, K., & Suzuki, H. (1992). "Measurement of cheek impedance by sound pressures and sound pressure in oral cavity and acceleration of vibrating cheek," J. Acoust. Soc. Jpn., 48, 621-628 (in Japanese).

[10] Flanagan, J. L., Ishizaka, K., and Shipley, K. L. (1975). "Synthesis of speech from a dynamic model of the vocal cords and vocal tract," BST J., 54, 485-506.