Visualization of Hypopharyngeal Cavities and Vocal-Tract Acoustic Modeling

Kiyoshi Honda
LPP, UMR7018 CNRS-Univ. Paris3
19, rue des Bernardins
75005 Paris, France
khonda@sannet.ne.jp

Tatsuya Kitamura
Konan University
8-91 Okamoto, Higashi-nada
Kobe 658-8501, Japan
t-kitamu@konan-u.ac.jp

Hironori Takemoto
NICT
2-2-2 Hikaridai, Seika-cho, Soraku-gun
Kyoto 619-0288, Japan
takemoto@nict.go.jp

Seiji Adachi
Fraunhofer Inst. Building Physics
Nobelstrasse 12
70569 Stuttgart, Germany

Parham Mokhtari
NICT
2-2-2 Hikaridai, Seika-cho, Soraku-gun
Kyoto 619-0288, Japan

Sayoko Takano
BAIC, ATR-Promotions
2-2-2 Hikaridai, Seika-cho, Soraku-gun
Kyoto 619-0288, Japan

Yukiko Nota
BAIC, ATR-Promotions
2-2-2 Hikaridai, Seika-cho, Soraku-gun
Kyoto 619-0288, Japan

Hiroyuki Hirata
NICT
2-2-2 Hikaridai, Seika-cho, Soraku-gun
Kyoto 619-0288, Japan

Ichiro Fujimoto
BAIC, ATR-Promotions
2-2-2 Hikaridai, Seika-cho, Soraku-gun
Kyoto 619-0288, Japan

Yasuhiro Shimada
BAIC, ATR-Promotions
2-2-2 Hikaridai, Seika-cho, Soraku-gun
Kyoto 619-0288, Japan

Shinobu Masaki
BAIC, ATR-Promotions
2-2-2 Hikaridai, Seika-cho, Soraku-gun
Kyoto 619-0288, Japan

Satoru Fujita
AI Inc.
1-7 Hikaridai, Seika-cho, Soraku-gun
Kyoto 619-0288, Japan

Jianwu Dang
JAIST
1-1 Asahidai, Nomi
Ishikawa 923-1291, Japan

Abstract

The hypopharyngeal cavities constitute a bottom part of the vocal tract near the larynx consisting of the supraglottic laryngeal cavity and the bilateral cavities of the piriform fossa. Visualization of those cavities with custom magnetic resonance imaging techniques reveals that during speech the laryngeal cavity takes a form of a long-neck flask and the piriform fossa a goblet of varying shapes: the former diminishes greatly in whispering and the latter disappears in deep inhalation. Our previous studies based on MRI have indicated that those cavities exert significant acoustic effects at higher frequency spectra. In this study, acoustic experiments with MRI-based mechanical vocal-tract models were conducted for male and female vocal tracts with the results that acoustic effects of those cavities determine frequency spectra above 2.5 kHz, giving rise to peaks and zeros. Those findings led us to propose an acoustic model of vowel production with three components: voice source, vocal-tract proper, and hypopharyngeal cavities. This model provides effective means in controlling voice quality and for expressing individual vocal characteristics.

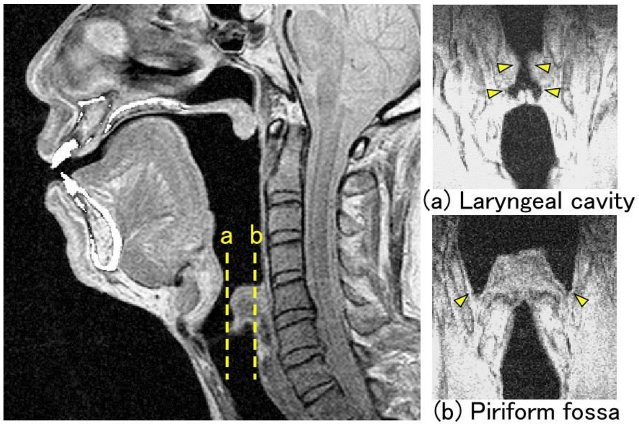
Keywords: hypopharynx, laryngeal cavity, piriform fossa, vocal tract, acoustic modeling.

1. Introduction

The vocal tract plays a role in producing natural speech with varying vowel qualities and voice qualities while maintaining relatively stable idiosyncratic features. The modern acoustic theory, e.g., the source-filter model, provides a full account of the vocal-tract function in controlling vowel qualities, while it is not successful in explaining other roles of the vocal tract in realizing human speech sounds. For example, if voice quality is due in large part to laryngeal source functions, then we question whether the wide range of voice quality in natural speech can be accounted for only by a small number of parameters of vocal-fold vibration. Similarly, if individual vocal characteristics arise in large part from idiosyncratic vocal-tract resonance patterns, then we wonder why speaker characteristics can be maintained with radically changing vocal-tract geometries in speech. Those questions lead us to ask if we may have overlooked certain critical elements of speech production processes, and certainly we know that we have failed to account for all the causal details of spectral envelopes of speech sounds. We have conducted our research by bearing those questions in mind to explore acoustic functions of realistic vocal tracts by means of magnetic resonance imaging (MRI), mainly focusing on the the geometry of the lower part of the vocal tract. Our previous studies have been successful to reveal those underexplored regions of the vocal tract, and we have proposed acoustic models of the piriform fossa [1] and the supraglottic laryngeal cavity [2, 3]. In this study, we summarize those findings toward a realistic acoustic model of vowel production that accounts for the details of speech sound control.

2. Visualization of Hypopharyngeal Cavities

The hypopharynx is a part of the vocal tract near the larynx that divides into three short tubes: the supraglottal laryngeal cavity and the bilateral cavities of the piriform fossa. Figure 1 shows those regions of interest obtained from static magnetic resonance imaging (MRI). In this figure, the laryngeal cavity is a basal segment of the vocal tract between the glottis and the bottom of the middle pharynx, which consists of bilateral spaces called the laryngeal ventricle and a narrow tube of the laryngeal vestibule above. The bilateral cavities of piriform fossa (often called the piriform sinuses) are a set of blind tubes of a conical shape formed by the aryepiglottic folds and the lateral hypopharyngeal wall, and they correspond to the entrance of the esophagus.



(a) Laryngeal cavity

(b) Piriform fossa

Figure 1. Midsagittal MRI of the vocal tract with denture (left), and coronal images of the hypopharynx showing (a) the laryngeal cavity and (b) bilateral cavities of the piriform fossa

Three-dimensional (3D) visualization of the hypopharynx with MRI requires special attentions due to its small size and natural movement of the larynx. The vertical dimension of both hypopharyngeal cavities is about 2 cm, which is much smaller than the organs commonly imaged by MRI. Also, natural movement of the larynx during quiet respiration or prolonged phonation often results in image blurring. Toward finer laryngeal image quality, we solved problems in imaging with two improvements. First, a custom surface coil was developed for laryngeal MRI by modifying the antenna of the radio-frequency (RF) receiver coil from that of an existing surface RF coil for a small joint. This modification allows laryngeal imaging with natural phonation ensuring a high signal-to-noise ratio in a small field of view. Second, an intermittent phonation-synchronized scan method was developed to obtain image data from phonatory phases during repetitive vowel productions with no data acquisition from respiratory phases. The combination of the two methods was successful in recording high-resolution laryngeal images with minimum motion artifacts.

3. Changes of Hypopharynx Geometry

The shape of the hypopharynx is relatively stable during natural vowel production regardless of vowel types, while the rest of the vocal tract region (vocal tract proper, in this article) shows large changes during speech [3]. Therefore, the hypopharynx is known to be one of the geometrical factors that cause individual vocal characteristics [4]. It is also known that the hypopharynx changes in shape when voice fundamental frequency is widely altered during vowels [5]. In laryngeal gestures other than natural phonation, however, the hypopharynx exhibits large variations in geometry. Figure 2 shows coronal MRI frames during humming, whispering, and deep respiration obtained from a male speaker with the larynx coil by means of phonation-synchronized MRI scans [6], which employs intermittent data acquisition during each repetition of humming, whispering, and inhalation with no data acquisition during other phases in the repetitions.

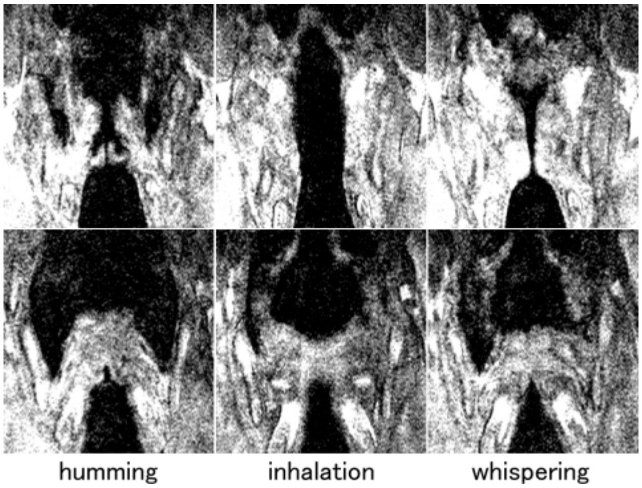


Figure 2. Coronal MRI for three laryngeal settings showing slices positioned for the laryngeal cavity (above) and for the piriform fossa (bottom).

In the figure, the upper and lower images in each modality correspond to two coronal scan planes passing though the laryngeal cavity and the piriform fossa, respectively. In humming, the hypopharynx takes about the same posture as in vowel production. In whispering, the laryngeal cavity forms a narrow passage with no obvious cavities of the laryngeal ventricles. Other frames show that the glottis is open only at the cartilagenous part. The piriform fossa is observed to be maintained, while it is slightly constricted along with a certain degree of constriction of the entire pharynx. These observations suggest that expiratory airflow is injected to the constricted laryngeal cavity via the posterior glottis to produce airflow noise for whispering. In deep respiration, the glottis opens wide, and the piriform fossa no longer exists in the images. This finding leads us to an interesting conjecture: the piriform fossa is not merely an empty cavity above the esophageal entrance but it offers a necessary space for the arytenoid cartilages to maximally abduct for a rapid intake of a large volume of air into the lung.

4. Previous Acoustic Studies on Hypopharynx

4.1 Piriform fossa

The piriform fossa has been known probably by early anatomists as a part of the hypopharynx, as it's name suggests. The "piriform" is a common anatomical term to describe body parts that resemble the shape of a pear. In the clinic, this cavity is visualized by upper esophageal fluoroscopy as an air cavity, where a small amount of liquid dwells momentarily before it is completely swallowed into the esophagus. In acoustic studies of the vocal tract, this cavity has been infrequently discussed partly due to its obscurity in lateral x-ray images. Chiba and Kajiyama [7], recognizing the cavity, described it as a short segment of a side branch in their vocal-tract diagrams, but they did not incorporate the cavity into the calculation of vocal-tract transfer functions. Fant [8] estimated acoustic effects of the piriform fossa on vowel formants, and subsequent studies by many researchers about the effect have discussed both vowel formants and higher frequency spectra (see Takemoto, et al. [2] for historical work).

Dang and Honda [1] conducted an MRI-based acoustic study of the piriform fossa using human subjects and mechanical models of the hypopharynx. When water is fed into the fossa, the spectral trough in the vicinity of 4-5 kHz disappears, showing evidence that the bilateral cavities of piriform fossa function as two side branches of the vocal tract emanating antiformants in that frequency region.

The bilateral cavities of the piriform fossa usually show left-right symmetry. When the cavities are modeled as two separate side branches, they should result in two zero-pole pairs, whereas it is often the case that only a single zero is observed in natural spectra in 4-5 kHz region. In models and humans in Dang and Honda [1], as water fills one fossa, a zero-pole pair shifts to the higher frequency. Then, another zero-pole pair appears in the similar frequency region. This observation may suggest that a zero-pole pair from one side of the piriform fossa annihilates the other zero-pole pair. Although acoustic interaction between the right and left cavities needs to be further examined, this conjecture accounts for frequently observed shapes of the spectral trough caused by the piriform fossa.

4.2 Supraglottic laryngeal cavity

The supraglottic laryngeal cavity (laryngeal cavity in this article) is a short axial segment of the vocal tract that consists of two anatomical regions: ventricles and vestibule of the larynx. The laryngeal ventricles (or Morgagni's hiatus) correspond to narrow bilateral spaces between the vocal and false vocal folds, and the laryngeal vestibule forms a narrow conduit from the ventricles to the opening at the middle pharynx. The shape of the laryngeal cavity resembles a Helmholtz resonator with a long neck. The ventricles form a very short segment in the vocal tract immediately above the glottis (about 2 mm long when measured along a vocal-tract midline). The entire laryngeal cavity has been considered as a simple straight tube (often called the "laryngeal tube") in acoustic modeling [7, 8], where the cavity widening at the ventricles has commonly been neglected due to their brevity.

Early acoustic studies of vowels have revealed that, in addition to the characteristic frequency regions (i.e., formants), there is a special resonance derived from the laryngeal tract [7], and this resonance is often reinforced in the western style of singing such as in opera [9], demonstrating a high-amplitude spectral peak called singer's formant.

Recently, we have examined the nature of the laryngeal cavity resonance by employing acoustic analyses and simulations based on vocal-tract MRI. A vocal-tract simulation with realistic shapes of the laryngeal cavity showed that both the ventricles and vestibule function independently to determine a peak frequency of the laryngeal-cavity resonance in agreement with the manner predicted by a model of Helmholtz resonator [2]. Further, supposing a tight closure of the glottal end, the laryngeal cavity is not involved in determining the lower formants of the vocal tract [2]. The laryngeal cavity produces a unique time function of resonance in synchrony with vocal-fold vibration, eliciting its resonance only at glottal closure phases, and the resonance damps out quickly at glottal opening [3]. The rise and fall of the laryngeal-cavity resonance within a glottal cycle can be observed in natural vowel production, offering a new acoustic analysis for measuring glottal closure period in ordinary spectrograms [3]. These studies suggest that voice quality derived from higher frequency resonance is finely controlled by laryngeal maneuvers to adjust the laryngeal-cavity resonance by means of cavity shape control for its frequency and of glottal closure action for its amplitude.

5. Modeling Hypopharyngeal Resonance

5.1 Method for data acquisition and solid modeling

The hypopharynx is not a uniaxial segment of the vocal tract, and acoustic modeling of the complex cavities requires special attentions. One way to achieve this task is acoustic experiment on mechanical models of realistic vocal tracts [10]. To do so, we used MRI data recorded during sustained production of five Japanese vowels to construct mechanical models of vocal tracts for a male and a female speaker. Volumetric MRI data for the entire vocal-tract regions were obtained from the two speakers using a phonation-synchronized scan method combined with a guide tone of harmonic sinusoidal waves [6] presented via a body-conduction sound transmitter [11]. 3D surfaces of the vocal tracts were extracted from the MRI data for the regions from the glottis to the perioral region without the nasal cavity. Then, the data were processed to have smoothed vocal-tract walls of 3-mm thickness and were fed to a rapid prototyping system to generate rigid-wall models of the vocal tracts, as shown in Figure 3. Acoustic recording experiments were conducted using white-noise signals from a mid-range horn driver unit (YL Model 551) to excite those models at the model's glottal end with different conditions to examine the effect of the piriform fossa and laryngeal cavity. Acoustic responses from the vocal tracts were measured with a digital sound recorder (Sony PCM-D50) at 48 kHz sampling in 16-bit resolution.

digital data mechanical model

Figure 3. Digital data of the vocal tract for vowel /a/ in front view (left) and a constructed solid model in side view (right)

5.2 Experiment on the piriform fossa

Experiments for examining the effect of the piriform fossa on vowel spectra were conducted with the same method used by Dang & Honda [1]. Prior to water injection into the cavities, the total volume of the bilateral cavities of the piriform fossa was measured for each model, which was about 2 ml for both male and female models regardless of vowel types. Acoustic recording was performed before and after water injection for each model with a glottal opening of 1.2 mm diameter (see below).

5.3 Experiment on the laryngeal cavity

Examination of the effect of the laryngeal cavity on vowel spectra was performed by simulating two conditions for open and closed glottis. The two conditions were realized by placing a metal plate of 2 mm thickness with a hole of different diameters as a septal wall between the model's glottal end and the opening of the horn driver unit. Closed and open glottis conditions for the male and female vocal tracts were simulated by holes of 1.2 mm and 4 mm diameter on the plate, respectively.

6. Results from Vocal Tract Experiments

Figure 4 shows the models' spectra for vowels from male vocal tracts. Blue and red lines correspond to data before and after water injection into the piriform fossa. Spectral differences are found mostly in the higher frequency region of 3.5 – 6 kHz. In the natural condition with no water in the fossa, the spectra show lower levels in that frequency region, typically with spectral zeros. In the water-filled condition, the spectra show one or more peaks in the same frequnency region, suggesting the appearance of the natural formants of the vocal tract with no piriform fossa. The fossa appears to produce large troughs at about 4-5 kHz and peaks at about 5-6 kHz in the vocal-tract transfer function, with a small frequency shift of vowel formants and a noticeable shift of the fifth formant toward the fourth.

Figure 5 shows spectra for female vocal tracts before and after water injection into the piriform fossa. Spectral differences are observed in the wider frequency region with a large vowel-to-vowel variation. Approximately the differences are seen in the two frequency regions. The spectral envelopes above 4 kHz are more attenuated in the natural condition than in the water-filled condition, similar to the results from the male vocal tracts. In the frequency region below 4 kHz, differences are mainly in the frequency domain, showing frequency shifts of the second and third formants, being lower in the natural condition.

Figure 6 shows the models' spectra from male vocal tracts in two conditions of the glottis. Blue and red lines correspond to closed and open glottis conditions, respectively. The major differences are seen in the spectral peak at about 3 kHz, which corresponds to the fourth formant in many vowels. In vowels /a/, /e/, and /o/, the spectral peak at 3 kHz in the closed glottis condition disappears in the spectra with the open condition. In vowels /i/ and /u/, the peak at 3 kHz in the closed condition reduces its peak level and increases its bandwidth in the open condition, indicating that the resonance of the laryngeal cavity disappears to exhibit a resonance of the vowel tract proper at the similar frequency. Frequency and amplitude changes of the first formant are commonly observed across vowels. Note that the elimination of the laryngeal-cavity resonance by opening the glottis does not significantly affect the level of the higher spectra above 4 kHz (actually in open glottis, the resonance mode is of a closed tube rather than of an open tube). This unique resonance pattern suggests that the laryngeal cavity derives an extra formant that is independent from the resonance of the vocal tract proper.

Figure 7 shows spectra for the female vocal tracts in the two conditions of the glottis. Spectral differences between the two conditions are seen in the wider frequency region in those data. Laryngeal-cavity resonance is found at 3.5 – 4 kHz with a broader bandwidth, and it appears to have an effect to increase the spectral level in the frequency region up to 4.5 kHz. Frequency and amplitude changes of the first formant are also observed across vowels, and there are no obvious differences in the higher frequency region between the two conditions.

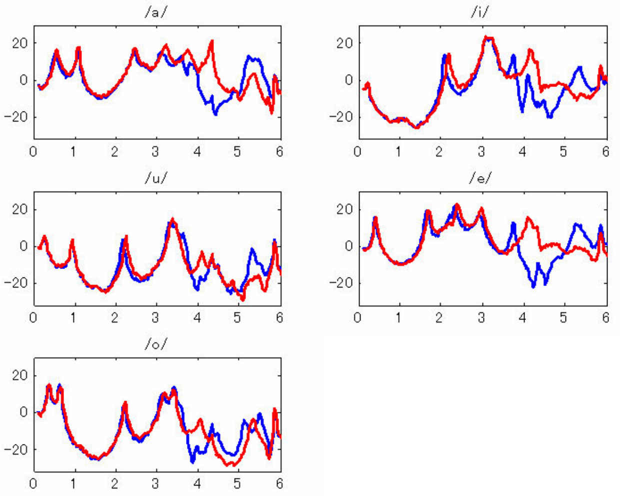


Figure 4. Vocal-tract transfer function for male models with empty fossa (blue) and filled fossa (red)

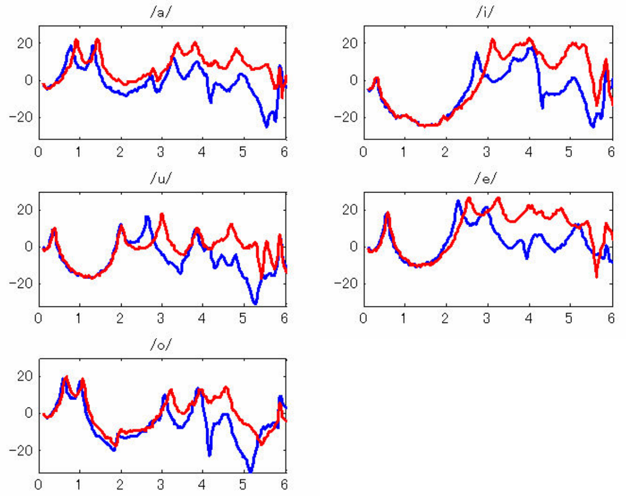


Figure 5. Vocal-tract transfer function for female models with empty fossa (blue) and filled fossa (red)

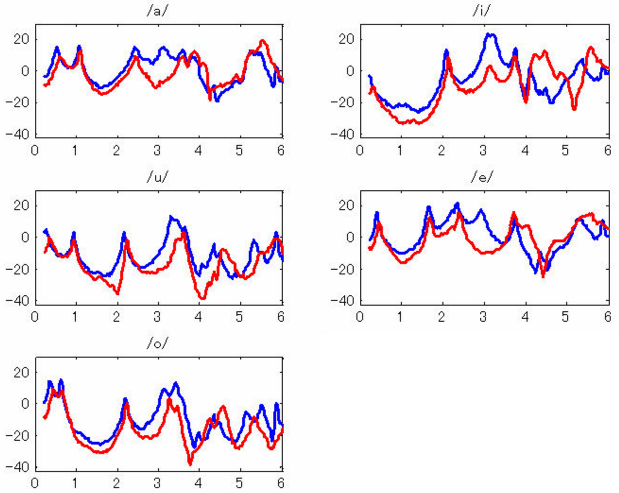


Figure 6. Vocal-tract transfer function for male models with closed glottis (blue) and open glottis (red)

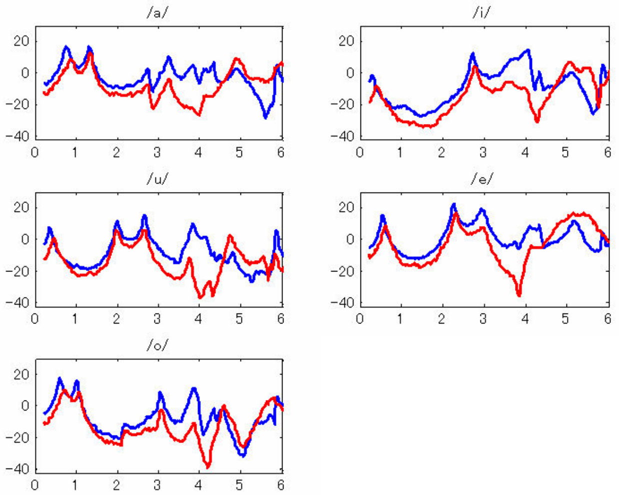


Figure 7. Vocal-tract transfer function for female models with closed glottis (blue) and open glottis (red)

7. Developing Vocal-tract Acoustic Model with Hypopharyngeal-cavity Resonance

The hypopharyngeal cavities function as a set of unique resonators in the vocal tract so as to modify vowel spectra remarkably in the higher frequency region. The supraglottic laryngeal cavity functions as an independent resonator functioning as a Helmholtz resonator to produce a resonance peak in the spectra at a 3 – 3.5 kHz region. This resonance is observed as an extra formant that elevates the spectral level around the region with no significant amplification in the higher frequencies. The bilateral cavities of the piriform fossa act as side branches of the vocal tract to emanate a zero-pole pair, attenuating sound energy in a 4-5 kHz region and amplifying the 5-6 kHz region. Those cavities also affect vowel formants to various degrees. These effects of hypopharyngeal-cavity resonance cannot be overlooked in the acoustic modeling of realistic vocal tracts, particularly when we consider supralaryngeal control of voice quality and causal factors of individual vocal characteristics.

Figure 8 shows a three-component model of vowel production that we propose based on our studies on the hypopharyngeal-cavity resonance as an extension of the source-filter account. This model employs voice source spectra, hypopharyngeal-cavity resonance, and resonance of the vocal tract proper as active components to control vowel spectra. The source sound determines fundamental frequency, amplitude, and spectral tilt. The hypopharyngeal-cavity resonance mainly modifies vowel spectra in the higher frequency region above 2.5 kHz with a resonance peak followed by a sharp down slope toward an antiresonance trough. The vocal tract proper determines the lower three formants.

The hypopharyngeal-cavity resonance is a component of particular importance in determining voice quality, since it involves a frequency region that is most sensitive to the human ears and is used for voice quality control in singing. For example, the accumulation of the third, fourth, and fifth formants is known to be used to produce a single peak of high amplitude called the singer's formant in the western style of singing. According to our model, this process is explained by the following three actions: (1) laryngeal maneuvers to lower the peak frequency of the laryngeal-cavity resonance either by widening the laryngeal vestibule or constricting the laryngeal vestibule, (2) extra-laryngeal forces to widen the piriform fossa by advancing the laryngeal structures so as to deepen the antiresonance trough and lower the fifth formant frequency, and (3) vocal-tract adjustment for raising the third formant frequency. Those actions result in a relatively constant amplitude level of the formant up to the singer's formant with a sharply declining spectral slope above the peak frequency of the singer's formant.

The hypopharyngeal-cavity resonance is also important as one of the causal factors of individual vocal characteristics, while it is an issue that still needs to be further investigated. It is generally recognized that individual differences in vocal-fold length and vocal-tract length are clear causal factors of speaker idiosyncracy. It has also been repeated in the history of acoustic studies of speech that there is a frequency region that signals speaker-specific sound characteristics. In addition to the fundamental frequency and formant distribution, the hypopharyngeal-cavity resonance contributes to realizing the idiosyncratic nature of voice by determining the higher frequency spectra, presumably together with a speaker-specific distribution of vowel formants including a certain spectral interaction between the hypopharyngeal-cavity resonance and the higher formants from the vocal tract proper.

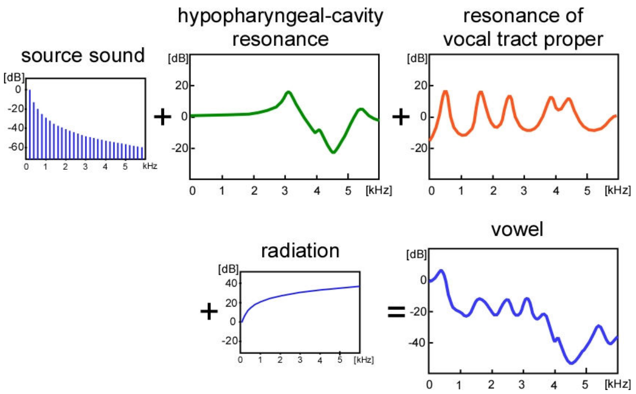


Figure 8. Acoustic model of vowel production with hypopharyngeal cavity resonance

8. Conclusion

We have conducted a long series of studies to reveal realistic acoustic functions of the vocal tract by spending many efforts for improving MRI image quality, re-examining vocal-tract anatomy, realizing mechanical models of vocal tracts, and conducting other works to support the proposed model of vocal-tract acoustics. In this article, we proposed an acoustic model of vowel production with hypopharyngeal-cavity resonance by summarizing our investigation. We emphasize that this model is beneficial over the previous source-filter model in accounting for variations of voice quality, idiosyncratic characteristics, and vowel formants. We believe that our model resolves many problems in speech technology discussed during the previous decades, such as unexpected failures of digital signal processing for estimating vocal-tract shapes inversely from speech signals, or unnatural speech quality of synthetic sounds from formant or articulatory synthesizers. We hope that further advancement of the model will contribute to expanding acoustic phonetics in the technical fields as well as in the basic fields of science and culture in relation to speech.

9. Acknowledgement

This research was conducted as part of “Research on Human Communication” with funding from the National Institute of Information and Communication Technology. This work is also supported by Kakenhi 21300071 and 21500184. The authors are grateful to Prof. Yutaka Ichinose (Information Science, Kyushu Sangyo Univ.) for his continuous support on our work at ATR Human Information Science Laboratories and to Prof. Eiji Yumoto (Otorhinolaryngology, Kumamoto Univ.) for suggesting MRI observation of the larynx during deep inhalation.

References

[1] J. Dang and K. Honda. “Acoustic characteristics of the human paranasal sinuses derived from transmission characteristic measurement and morphological observation,” *J. Acoust. Soc. Am.*, vol. 100, pp. 3374-3383, 1996.

[2] H. Takemoto, S. Adachi, T. Kitamura, P. Mokhtari, and K. Honda. “Acoustic roles of the laryngeal cavity in vocal tract resonance,” *J. Acoust. Soc. Am.*, vol. 120, pp. 2228-2238, 2006.

[3] T. Kitamura, H. Takemoto, S. Adachi, P. Mokhtari, and K. Honda. “Cyclicity of laryngeal cavity resonance due to vocal fold vibration,” *J. Acoust. Soc. Am.*, vol. 120, pp. 2239-2249, 2006.

[4] T. Kitamura, K. Honda, and H. Takemoto. “Individual variation of the hypopharyngeal cavities and its acoustic effects,” *Acoust. Sci. & Tech.*, vol. 26, pp. 16-26, 2005.

[5] H. Takemoto, T. Kitamura, K. Honda, and S. Masaki. “Deformation of the hypopharyngeal cavities due to F0 changes and its acoustic effects,” *Acoust. Sci. & Tech.*, vol. 29, pp. 300-303, 2008.

[6] S. Takano, K. Kinoshita, and K. Honda. “Measurement of cricothyroid articulation using high-resolution MRI and 3D pattern matching,” *Acta Acustica united with Acustica*, vol. 92, pp. 725-730, 2006.

[7] T. Chiba, and M. Kajiyama. *The Vowel – Its Nature and Structure*, Tokyo: Tokyo-Kaiseikan, 1942.

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

[9] J. Sundberg. “Formant structure and articulation of spoken and sung vowel,” Folia Phoniatr., vol. 22, pp. 28-48, 1970.

[10] S. Fujita and K. Honda. “An experimental study of acoustic characteristics of hypopharyngeal cavities using vocal tract solid models,” *Acoust. Sci. & Tech.*, vol. 26, pp. 353-357, 2005.

[11] Y. Nota, T. Kitamura, K. Honda, H. Takemoto, H. Hirata, Y. Shimada, and I. Fujimoto. “A bone-conduction system for auditory stimulation in MRI,” *Acoust. Sci. & Tech.*, vol. 28, pp. 33-38, 2007.