ISCA Archive
http://www.isca-speech.org/archive

Speech Prosody 2004
Nara, Japan
March 23-26, 2004

Acoustical Interpretation of Certain Laryngeal Settings Using a Physical Model

Ken-Ichi Sakakibara

NTT Communication Science Laboratories
NTT Corporation
`kis@brl.ntt.co.jp`

Hiroshi Imagawa

Department of Speech Physiology
The University of Tokyo
`imagawa@m.u-tokyo.ac.jp`

10.21437/SpeechProsody.2004-147

Abstract

This paper examines the acoustical effects of certain different laryngeal settings regulated by the shape of the larynx tube and hypopharynx on vowels. These effects are described using a two-by-two mass model, which is realized as two-layer two-mass models. The five Japanese vowels in different laryngeal settings were synthesized using the two-by-two mass model, with the area function of the vocal tract obtained by MRI. When the supraglottal structure of the larynx is constricted, the effects of the laryngeal ventricle resonance appear, and they depend on the extent of the constriction. In such cases, the inverse-filtered voices are not equal to the laryngeal sources.

1. Introduction

In the framework of source-filter theory, for sustained vowel-like phonation, voice quality may be mainly related to the laryngeal source or laryngeal setting, although supralaryngeal settings, such as the velopharyngeal setting, also influence voice quality.

Various styles of singing show considerable variations in voice quality and associated laryngeal settings. In operatic singing, the vocal fold vibratory patterns mainly regulate its voice quality. However, in operatic singing, the expansion of the hypopharyngeal space and the lowering of the larynx height have been observed, and these phenomena influence the voice quality. The hypopharyngeal space expansion induces a special formant cluster, the so-called singer's formant [13]. In throat singing, the supraglottal structure of the larynx strongly constricts and the ventricular folds vibrate [3, 11]. The constriction and vibration of the ventricular folds play an important role in generating special glottal sources of throat singing [10, 11].

In speech, the vocal fold vibratory patterns and glottal chink are emphasized for the regulation of voice quality, as characterized by breathiness, pressedness, register, roughness, and so on. However, in the laryngeal adjustments associated with some aspects of voice quality, such as harshness and some aspects of roughness, the constriction of the supraglottal structure plays an important role [7]. Therefore, the acoustical effects of the shape of the larynx tube must be taken into consideration for a better understanding of the regulation mechanism of the voice quality.

In this paper, we investigate the acoustical effects of the laryngeal settings using a two-by-two mass model, which is realized as a two-layer two-mass models and permits ventricular fold vibrations (Fig. 2) [4, 10, 11]. This physical model connects the larynx and vocal tract in terms of acoustical coupling, and the influences of the vocal tract on the voice source is take into consideration [5, 9]. Using this model, we synthesize the Japanese five vowels in different laryngeal settings and investigate the acoustical characteristics of sounds filtered by the vocal tract. We also estimate the laryngeal source of the synthesized sound using inverse-filtering and discuss problems with the inverse-filter method for different laryngeal settings.

2. Anatomical overview of the larynx

The larynx vertically extends from the laryngeal entrance to the inferior border of the cricoid cartilages. The entrance of the larynx is surrounded by the epiglottis, the aryepiglottic folds, and the mucosa between the arytenoids (Fig. 1).

The ventricular folds (false vocal folds) are a pair of flaccid folds located above the laryngeal ventricle. Whereas the vocal folds have a mechanism that changes their stiffness, thickness, and longitude, the ventricular folds are incapable of becoming tense, since they contain very few muscle fibres. They are abducted and adducted by the action of certain laryngeal muscles. In normal phonation, seemingly, they do not vibrate.

The laryngeal ventricle is a space between the vocal and ventricular folds. The aryepiglottic folds are the superior edges of the quadrangular membranes linking the arytenoids to the epiglottis. The larynx is separated into three parts vertically: supraglottal, glottal, and subglottal regions. The supraglottal region is also called the larynx tube. The laryngeal vestibule is the region between the ventricular folds and the laryngeal entrance. The region just above or around the larynx is called the hypopharynx. The pyriform sinuses consists of two cavities in the bottom of the hypopharynx, that is, just above the esophageal entrance.

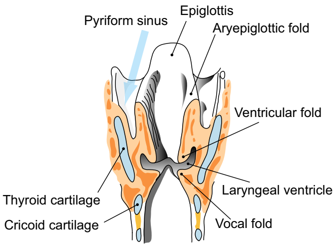


Figure 1: *Coronal view of the larynx, as seen from behind.*

3. Physical model

The mechanism of the physical model used here is depicted in Fig. 2. The vocal and ventricular folds are described by the 2×2-mass model, which represents the ventricular folds in a self-oscillating model as well as the vocal folds [4, 10]. This model was obtained by improving the two-mass model [5]. Fig. 3 shows the equivalent circuit of our proposed model.

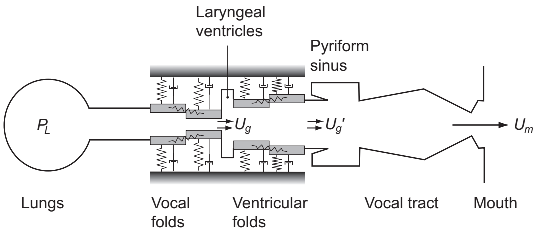


Figure 2: *Physical model for synthesis of singing voices.*

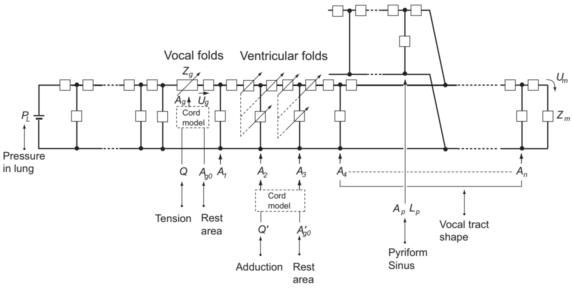


Figure 3: *Equivalent circuit of the synthesis model.*

The vocal tract is represented as a lossy transmission line [8] of 43 cylindrical sections with length 0.4 cm. Hence, the total length of the vocal tract is $17.2\,\mathrm{cm}$.

The bilateral pyriform sinuses are assumed to be symmetric and, therefore implemented as one cavity. Based on MRI data [1], we represent each pyriform sinus as a cone whose depth is $2.0\,\mathrm{cm}$ and volume is $1.5\,\mathrm{cm}^2$. We use a value of 0.75 for the end correction coefficient. The result of a preliminary experiment showed that the acoustic characteristics are not significantly different between one-cavity and two-cavity cases.

The effect of the nasal cavity is neglected for simplicity because the range of nasality for each vowel is still controversial and individual differences are large. The cross-sectional areas of the subglottal system were determined based on the anatomical data in [6].

4. Parameter setting

We set the length of each part of the vocal tract as shown in Fig. 4. The length of the larynx tube is $2.4\,\mathrm{cm}$. The larynx tube includes the ventricular folds sections ($0.8\,\mathrm{cm}$), laryngeal ventricle section ($0.4\,\mathrm{cm}$), and laryngeal vestibule ($1.2\,\mathrm{cm}$). We also set the default length of the hypopharynx to $1.6\,\mathrm{cm}$. The physical properties of the vocal folds were set to the normal values in [5]. The physical properties of the ventricular folds were set based on those of the vocal folds [10]. The areas of the 32 sections above the hypopharynx vary depending on the vowel. The area functions for the five Japanese vowels, /a/, /i/, /u/, /e/, /o/, were determined from MRI data [15].

The five different laryngeal settings are (1) normal, (2) operatic singing, (3) pressed, (4) drone of throat singing, (5) kargyraa of throat singing. The area of the laryngeal ventricle was set to $1.5\,\mathrm{cm}^2$ for all settings. For the normal setting, the areas of the ventricular folds and laryngeal vestibule sections were set to $0.8, 1.0\,\mathrm{cm}^2$, respectively. These values were determined from the anatomical observation of the human larynx, and show good agreement with the data in [2, 14]. We set the subglottal pressure $P_L = 5\ \mathrm{cm\,H_2O}$ in the normal setting (Fig. 4).

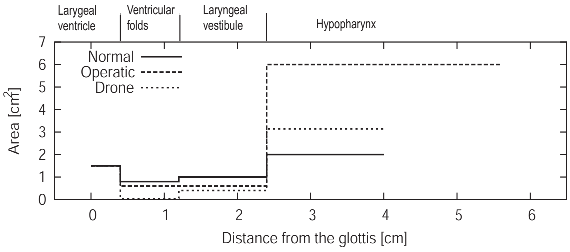


Figure 4: *Larynx tube and hypopharyngeal settings.*

For operatic singing, the areas of the ventricular folds and laryngeal vestibule sections were uniformally set to $0.6\,\mathrm{cm}^2$. In operatic singing, the larynx is lowered and the spaces of the hypopharyngeal and pyriform sinuses are expanded. Therefore, the sections of the hypopharynx are extended $1.6\,\mathrm{cm}$ [12, 16] and the hypopharynx area was set to $6.0\,\mathrm{cm}^2$. The volume of the pyriform sinuses was also expanded to 1.5 times that of the normal setting. We set the subglottal pressure $P_L = 5\ \mathrm{cm\,H_2O}$ in operatic singing.

For drone, kargyraa, and pressed settings, the areas of the laryngeal vestibule and hypopharynx were set to 0.4

and $3.14\,\mathrm{cm}^2$ respectively. The differences among three settings are regulated by the extent of the ventricular fold adduction: Q', i.e., we set the rest area of the ventricular fold section $A'_{g0} = 0.04/Q'\,\mathrm{cm}^2$ for each [10]. In drone setting with $Q' = 1.0$, the ventricular folds vibrate with the same period as the vocal fold vibration. In the kargyraa setting with $Q' = 0.55$, the ventricular folds vibrate with double-period of the vocal fold vibration. In the pressed setting with $Q' = 0.1$, the ventricular fold vibration is small enough to neglect. In some naming schemes for voice quality, this pressed setting is called tensed rather than pressed. We set the subglottal pressure $P_L = 5\ \mathrm{cm\,H_2O}$ for pressed and $20\,\mathrm{cm\,H_2O}$ for drone and kargyraa.

In our previous work [4], we observed formant frequencies, which are regulated by the shapes of the hypopharynx and larynx tube, on /e/. The results are summarized as follows: decreasing the larynx tube volume moves F_2, F_3, F_4 close to F_2. Increasing the volume of the hypopharynx generates the formant cluster of F_3, F_4, F_5 (singer's formant). Increasing the volume of the laryngeal ventricle moves F_3, F_4 lower.

5. Synthesized vowels

Figs. 5–9 show LPC spectral envelopes of the synthesized five Japanese vowels in the five different laryngeal settings (LPC order 26, 22.05 kHz sf).

The normal setting shows the largest F_1 power among the five settings. The influence of the zero by the pyriform sinuses, which should appear at around 4.5 kHz, is seemingly smallest at the normal setting. In operatic singing, the singer's formant appears slightly above 3 kHz commonly among the five vowels. As pointed out in previous papers [4, 11, 13], the generation of the singer's formant is due to the laryngeal ventricle resonance, supported by the hypopharynx, which is large enough in comparison with the area of the laryngeal vestibule. In the drone voice, compared with other settings, F_2 is relatively sharp in most vowels except /e/.

6. Laryngeal and glottal sources

Fig. 10 shows the spectral envelope of the inverse-filtered source (left), synthesized laryngeal flow u'_g (center), and synthesized glottal flow u_g (right) of /e/ at the five laryngeal settings. The inverse-filtering was processed directly using $F_1, \ldots, F_5$ in LPC analysis.

For normal and pressed, the spectra of the inverse-filtered voices are similar to that of the glottal sources. However, in operatic singing, in the range of around 4.5 kHz, the inverse-filtered voice has larger power than the glottal source. This phenomenon presumably arises because the zero of the pyriform sinuses lowered F_3, F_4, F_5. The dips at around 4 kHz are commonly observed at all settings. These dips are caused by the zero of the pyriform sinuses. In drone and kargyraa, the inverse-filtered voices are different from both laryngeal and glottal sources. These differences may be due to the non-linearity caused by the ventricular fold vibration in the larynx and the laryngeal ventricle resonance. Fig. 11 shows three sources in drone.

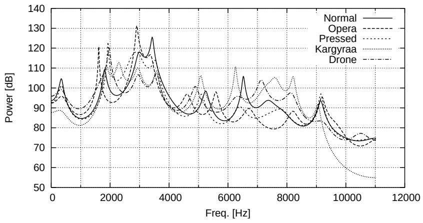


Figure 5: *Spectral envelopes of /i/.*

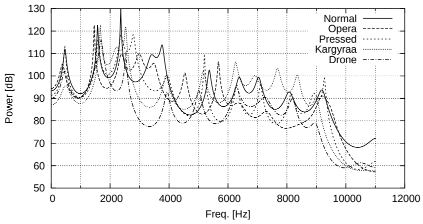


Figure 6: *Spectral envelopes of /e/.*

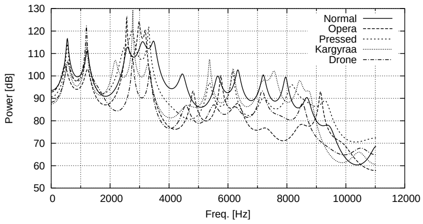


Figure 7: *Spectral envelopes of /a/.*

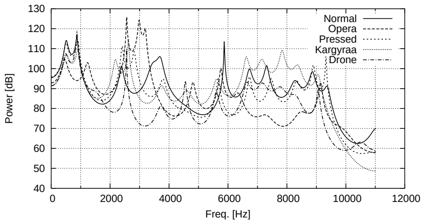


Figure 8: *Spectral envelopes of /o/.*

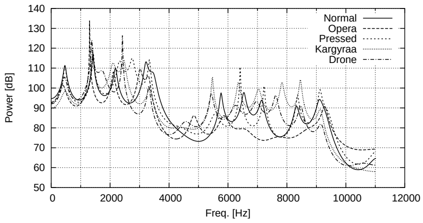
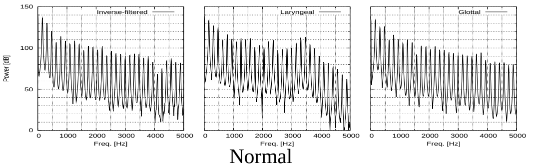
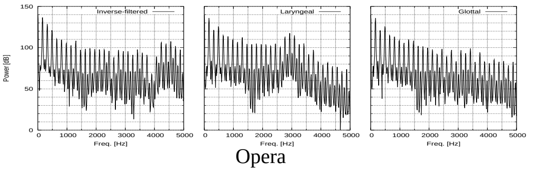


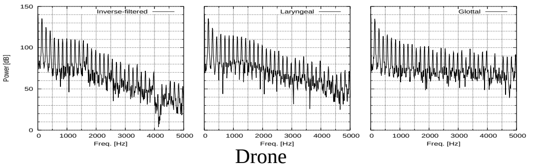
Figure 9: *Spectral envelopes of /u/.*



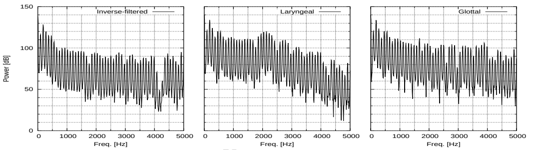
Normal



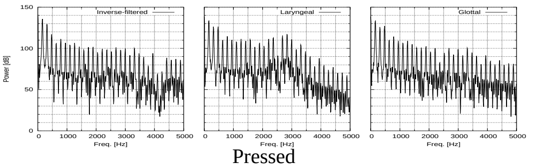
Opera



Drone



Kargyraa



Pressed

Figure 10: *Spectra of inverse-filtered (left), laryngeal (center), and glottal (right) sources in different phonations.*

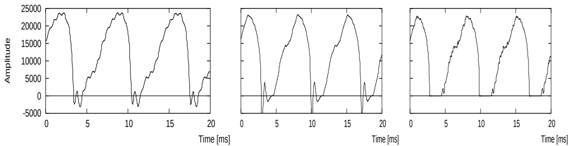


Figure 11: *Inverse-filtered (left), laryngeal (center), and glottal (right) sources in drone.*

7. Discussions and conclusions

The shapes of the hypopharynx and larynx tube have non-negligible influences on the spectral shapes of the sound and sources. The constriction of the supraglottal structure of the larynx regulates these shapes and often correlates to emotional states and other important non-linguistic attributes.

The extent of the constriction is not reflected on the glottal source, but on the laryngeal source. The inverse-filtering is considered to be one of the most powerful methods to extract the laryngeal source. However, the inverse-filtering assumes an all-pole model of the vocal tract and monotonous decreasing of the spectral envelope of the source. It therefore works well on the normal setting, but may not work on special laryngeal settings, such as harsh, or with supraglottal vibration, because the laryngeal ventricle resonance is included in the laryngeal source.

The simulation here may be useful in establishing an appropriate laryngeal flow model, such as in [11]. We will also improve the inverse-filtering method so that it can be used for effective analysis of various aspects of voice quality.

Acknowledgments

We thank Kiyoshi Honda, Parham Mokhtari, Seiji Niimi, and Niro Tayama for their helpful discussions.

8. References

[1] J. Dang and K. Honda. Acoustic characteristics of the piriform fossa in models and humans. *J. Acoust. Soc. Am.*, 101(1):456–465, 1997.

[2] G. Fant and M. Båvegård. Parametric model of VT area functions: vowels and consonants. *KTH TMH-QPSR*, 1/1997:1–21, 1997.

[3] L. Fuks, B. Hammarberg, and J. Sundberg. A self-sustained vocal-ventricular phonation mode: acoustical, aerodynamic and glottographic evidences. *KTH TMH-QPSR*, 3/1998:49–59, 1998.

[4] H. Imagawa, K.-I. Sakakibara, N. Tayama, and S. Niimi. The effect of the hypophyaryngeal and supra-glottic shapes for the singing voice. *Proc. Stockholm Music Acoustics Conference 03*, II:471–474, 2003.

[5] K. Ishizaka and J. L. Flanagan. Synthesis of voiced sounds from a two-mass model of the vocal cords. *Bell System Tech . J.*, 51(6):1233–1268, 1972.

[6] K. Ishizaka, M. Matsudaira, and T. Kaneko. Input acoustic-impedance measurement of the subglottal system. *J. Acoust. Soc. Am.*, 60(1):190–197, 1976.

[7] J. Lindqvist-Gauffin. A descriptive model of laryngeal articulation in speech. *KTH STL-QPSR*, 2–3/1972:1–9, 1972.

[8] S. Maeda. A digital simulation method of the vocal-tract system. *Speech Communication*, 1:199-229, 1982.

[9] M. Rothenberg. Acoustic interaction between the glottal source and the vocal tract. In *Vocal fold physiology*, K. N. Stevens and M. Hirano (eds.), 305–328, Univ. Tokyo Press., 1981.

[10] K.-I. Sakakibara, H. Imagawa, S. Niimi, and N. Osaka. Synthesis of the laryngeal source of throat singing using a 2×2-mass model. *Proc. of International Computer Music Conference 2002*, 5–8, 2002.

[11] K.-I. Sakakibara, T. Konishi, K. Kondo, E. Z. Murano, M. Kumada, H. Imagawa, and S. Niimi. Vocal fold and false vocal fold vibrations and synthesis of khöömei. *Proc. of International Computer Music Conference 2001*, 135–138, 2001.

[12] T. Shipp and K. Izdebski. Vocal frequency and vertical larynx positioning by singers and nonsingers. *J. Acoust. Soc. Am.*, 58(5):1104–1106, 1975.

[13] J. Sundberg. Articulatory interpretation of the "singing formant". *J. Acoust. Soc. Am.*, 55(4): 838–844, 1974.

[14] H. Takemoto, K. Honda, S. Masaki, Y. Shimada, and I. Fujimoto. Modeling of the inferior part of the vocal tract based on analysis of 3D cine-MRI data. *Proc. 2003 Autumn Meeting Acoust. Soc. Jpn.*, I:281–283, in Japanese, 2003.

[15] M. Tiede. MRI database and extraction of 3D vocal tract shape. Technical reporrt, ATR, 1999.

[16] S. Wang. Singing voice: bright timbre, singer's formants and larynx positions. *Proc. Stockholm Music Acoustics Conf. 83*, 313–322, 1983.