Forum Acusticum 2005

A many-parameter model of laryngeal flow with ventricular resonance and supraglottal vibration

Ken-Ichi Sakakibara
e-mail: **kis@brl.ntt.co.jp**
NTT Communication Science Laboratories, NTT Corporation
3-1, Morinosato Wakamiya, Atsugi-shi, 243-0198, Japan

Hiroshi Imagawa
e-mail: **imagawa@m.u-tokyo.ac.jp**
Department of Otolaryngology, The University of Tokyo
7-3-1, Hongou, Bunkyo-ku, Tokyo, 113-0033, Japan

A laryngeal flow model with many parameters is described. This model consists of a glottal flow model, laryngeal ventricle resonator, and supraglottal modulator. The model is capable of representing various laryngeal airflows even when they are accompanied by supraglottal constriction and vibrations, such as found in throat singing, growl, and harsh voices. The proposed model contains several parameters related to a supraglottal modulation derived from the supraglottal vibration as well as those related to a glottal flow. The parameters of the ventricular modulation is set depending on vocal styles. For example, in subharmonic ventricular vibration, the parameter of modulation frequency is set to an integer multiple as the frequency of the vocal fold vibration. The laryngeal ventricle resonance is implemented as a time-variant one-pole filter in the laryngeal flow model. The proposed model is compared with laryngeal flow simulated by using a physical modeling, which is a self-oscillating mass-spring model for vocal and ventricle folds (2×2-mass model). It is shown that the proposed model provides a good fit to the estimated laryngeal flow in vocal-ventricular phonation.

1 Introduction

The glottal source model has been widely studied because the glottal flow is expected to play an important role in determining voice quality. Several different glottal source models, such as the Rosenberg model (R-model) [10] and the Liljencrants-Fant model (LF-model) [4], have been proposed and used for analyses and syntheses in the framework of the source-filter model. In particular, the LF-model has become a reference for glottal-source analysis related to various voice qualities, such as modal, falsetto, vocal fry, pressed, and breathy voices, in both of speech and singing [1, 2, 8].

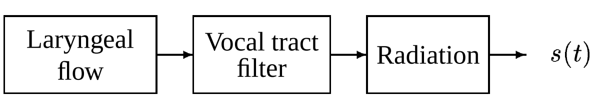

Figure 1: Source-filter model with laryngeal source, vocal-tract filter and radiation.

All glottal source models assume that the glottal voice source is determined by vocal fold vibratory patterns and aim at controlling voice quality by changing the parameters related to the vocal fold vibratory pattern.

In some singing styles, besides the vocal folds, the supraglottal structure in the larynx, including the ventricular folds, aryepiglottic folds, and laryngeal ventricle, often plays an important role in regulating voice quality. In throat singing, such as Khöömei in Tyva, Khöömij in Mongolia, and Kai in Altai, the the ventricular folds (also referred to as the false vocal fold) vibrate and contribute to the generation of its unique timbre [5, 7, 14]. Furthermore, among traditional music styles in Asia, a pressed-type singing techniques is widely employed. Japanese Min-yoh, Korean Pansori, and Chinese Beijin Opera are its representatives. For example, in Min-yoh, the constriction of the supraglottic structure is also observed, though the ventricular folds do not vibrate [6]. In growl, which is widely found in Blues, Gospel, and Jazz, the aryepiglottic folds vibrate, as well as the vocal folds [11].

In speech, the vibration of the supraglottal structure in the larynx, including the aryepiglottic and ventricular folds, contributes to phonemic contrasts in some languages [3]. Moreover, in harsh-type voices, in terms of Laver's voice quality description, laryngeal sources obtained by inverse-filtering are mostly subharmonic [9]. While the subharmonic voice sources in some of the creak voices, which are also referred to as vocal fry, arise from vocal fold vibratory patterns, the subharmonic source of the harsh-type voice is presumably the vibration of the supraglottal structure in the larynx, such as the aryepiglottic folds.

Therefore, for synthesis of various styles of singing voices, besides the vocal fold vibration, the effects of vibration of the supraglottal structure, including the ventricular and aryepiglottic folds, and the resonance of the laryngeal ventricle must be considered.

In this paper, we propose a many-parameter laryngeal model based on glottal flow, laryngeal ventricle resonance, and the modulation of the ventricular fold vibration. As the laryngeal source for the source-filter synthesis, the proposed model is able to manipulate various voice qualities.

2 Laryngeal sources

Here, we use *laryngeal flow (source)* $u_l(t)$ to mean the airflow through the whole larynx or epilarynx, and *glottal flow (source)* $u_g(t)$ to mean the airflow through the glottis.

Ventricular fold vibration is observed in various types of phonation in singing and in voice disorders. In throat singing, both drone and kargyraa voice phonations are always accompanied by ventricular fold vibrations, as well as by vocal fold vibrations. In the drone voice, the ventricular folds vibrate in the same period as the vocal folds, and in the kargyraa voice, the ventricular folds vibrate in an integer multiple (usually double or triple) period of the vocal folds [5, 7, 14].

2.1 Laryngeal flow by inverse-filtering

The laryngeal flows of drone and kargyraa for two different male throat singers are shown in Figs. 2, and 3. These flows were obtained from recorded sounds using an inverse-filter analysis. We marked five poles on the spectrum in the range from 0 to 5 kHz, constructed the inverse-filter, and manually adjusted it to make the resulting flow smooth.

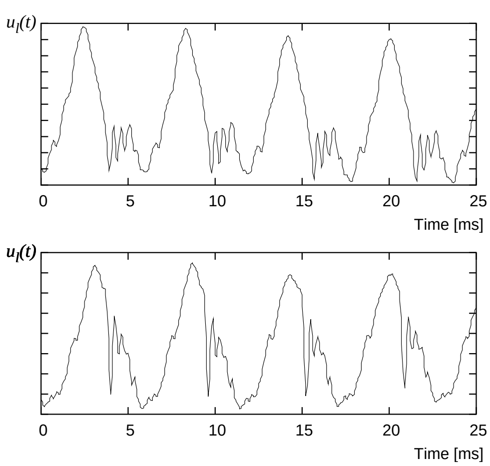

Figure 2: Laryngeal airflow of drone voice obtained by inverse-filtering. Top: Singer A. Bottom: Singer B.

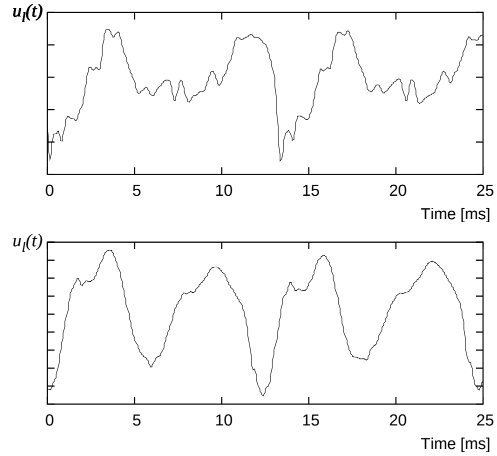

Figure 3: Laryngeal airflow of kargyraa voice obtained by inverse-filtering. Top: Singer A. Bottom: Singer B.

2.2 Simulation using a physical modeling

For simulating drone and karygraa phonations using a physical model, a certain vibratory mechanism is needed for the ventricular folds in the model.

The 2×2-mass is a two-tiered self-oscillating mass-spring model obtained by improving the Ishizaka-Flanagan's two-mass model. In the 2×2-mass model, the ventricular folds are described by a self-oscillating model as are the vocal folds [12, 14] (Fig. 4).

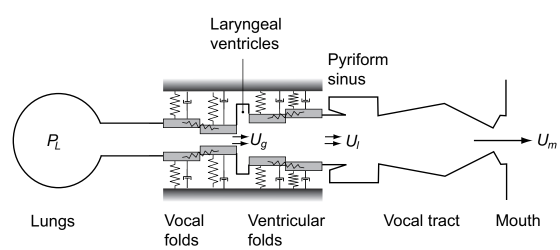

Figure 4: Physical model for synthesis of singing voices.

The ventricular folds are not equipped with a mechanism to change their physical properties, but they can be adducted by the action of certain laryngeal muscles. Their physical properties, such as mass and stiffness, and how those properties are changed by the adduction are still unclear. To take into account the changing shapes of the ventricular folds, we introduce an *adduction parameter* into the model, as one possible parameterization. Along with a narrowing of the rest area between the ventricular folds in the 2×2-mass model, various phonations—pressed without distinguishable ventricular vibration (e.g., Min-yoh, Japanese singing, [6]), kargyraa (triple- and double-periodic), and drone—appear consecutively. These results are supported by physiological observations [12]. The simulated laryngeal flow includes the laryngeal ventricle resonance.

In Fig. 5, synthesized laryngeal flows (top) and the vocal-ventricular vibratory patterns (bottom) at two adduction parameter values are shown. In the bottom panels, the thick solid line, thick dotted line, thin solid line, and thin dotted line show the horizontal displacements of the lower lip of the vocal fold, upper lip of the vocal fold, lower lip of the ventricular fold, upper lip of the ventricular fold, respectively.

In the panels on the left, the area between ventricular folds is initially set to $0.04\ \mathrm{cm}^2$. The ventricular adduction is considerably strong, and the initial setting was interpreted from the physiological observation of the drone voice. The simulated vocal-ventricular vibratory pattern is very similar to the vibratory pattern observed using high-speed imaging and EGG [14]. Namely, the frequency of the ventricular fold vibration is equal to that of the vocal fold vibration, and a phase difference of vocal and ventricular folds is about $\pi/4$. The synthesized laryngeal voice is also very similar to the laryngeal voice of drone obtained by inverse-filtering (Fig. 2).

In the panels on the right, a ventricular adduction is strong but looser that that in the drone voice setting. The are between ventricular folds was initially set to $0.08\ \mathrm{cm}^2$. The initial setting was also interpreted from the physiological observation of the kargyraa voice. The simulated vocal-ventricular vibratory pattern is very similar to the vibratory pattern observed using high-speed imaging and EGG [14]. Namely, the frequency of the ventricular fold vibration is half that of the vocal fold vibration. The synthesized laryngeal voice is also very similar to the laryngeal voice of kargyraa obtained by inverse-filtering (Fig. 3).

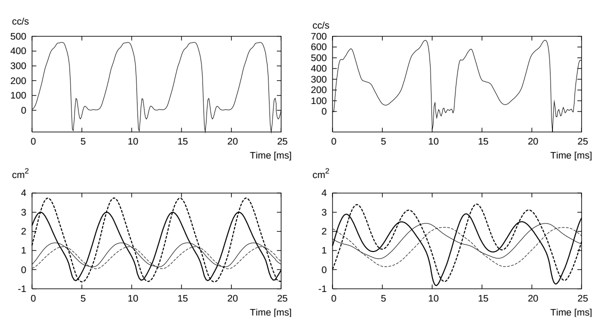

Figure 5: Synthesized laryngeal flows using the 2×2-mass model. Left: Simulated drone voice. Right: Simulated double-periodic kargyraa voice. Laryngeal flow (top) and the vocal and ventricular fold vibrations (bottom).

3 Laryngeal flow model

Based on the above results, a new laryngeal flow model that makes it possible to synthesize various voice timbres is proposed. Here, the model is obtained through two stages. In the first stage, we construct a ventricular modulation model (VTF-modulation model), and in the second, a VTF-modulation and larynx tube resonance model (VTF-modulation+LT-resonance model).

3.1 VTF-modulation model

The block diagram of a VTF-modulation model is shown in Fig. 6.

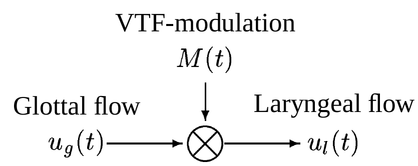

Figure 6: Block diagram of the VTF-modulation model

The VTF-modulation model of the laryngeal flow $u_l(t)$ is simply defined as follows:

$$u_l(t) := M(t)u_g(t) \tag{1}$$

The vibratory patterns of the ventricular folds were observed using a high-speed imaging and seem to not be exactly a sine-shape [7, 14]. However, here we define the VTF-modulation function $M(t)$ by multiplication by constant M of the false glottal area function A_g'. We also define A_g' as an asymmetric cosine function:

$$M(t) := MA_g'(t) \tag{2}$$

$$A_g'(t) := \begin{cases} -\alpha' \cos\left(\pi \frac{t}{T_p{}'} - \theta'\right) + A_{g_0}{}', \\ \qquad\qquad \text{if} \quad 0 \le t \le T_p{}' \\ \alpha' \cos\left(\pi \frac{t - T_p{}'}{T_0{}' - T_p{}'} - \theta'\right) + A_{g_0}{}', \\ \qquad\qquad \text{if} \quad T_p{}' \le t \le T_0{}' \end{cases} \tag{3}$$

where α' represents the amplitude of the ventricular fold vibration, $A_{g_0}{}'$ the area between the ventricular folds at rest, $T_0{}'$ the period of the ventricular fold vibration, $T_p{}'$ the time at which the ventricular fold is open maximally, and θ' the phase difference of the ventricular fold vibration from vocal fold vibration. All of these are positive real numbers. Physiological observations and simulation using the 2×2-mass model suggest that the periods of the vocal and ventricular fold vibration satisfy

$$T'_0 = nT_0 \tag{4}$$

where n is a positive integer.

The speed quotient $S_q{}'$ for the ventricular fold modulation is defined as

$$S_q{}' := \frac{T_p{}'}{T_0{}' - T_p{}'} \tag{5}$$

A R-model [10] is used for the glottal flow. The R-model used here [the type B in [10]] is described as follows:

$$u_g(t) = \begin{cases} \alpha \left(3 \left(\frac{t}{T_p} \right)^2 - 2 \left(\frac{t}{T_p} \right)^3 \right), \\ \qquad \text{if} \quad 0 \le t \le T_p \\ \alpha \left(1 - \left(\frac{t-T_p}{T_p} \right)^2 \right), \\ \qquad \text{if} \quad T_p \le t \le T_p + T_n \le T_0 \end{cases} \tag{6}$$

where α is amplitude, T_p opening time, T_n closing time, and T_0 the period. All of these variables are positive real numbers. The open quotient O_q is written as $(T_p + T_n)/T_0$.

Double-periodic subharmonic laryngeal flow is represented by using a period changing parameter T_q. As seen in Fig. 5, in subharmonic vibration of the vocal and ventricular folds, the second flow $u_{g_2}(t)$ starts during the closing phase of the first flow $u_{g_1}(t)$. The starting time of the second flow was set to $T_q T_0$. For periodic normal glottal flow, $T_q = 1$. In one double period, glottal flow $u_g(t)$ is defined as

$$u_g(t) = \max(u_{g_1}(t), u_{g_2}(t - T_q T_0)) \tag{7}$$

In the case of n-multiple periodic vibrations (Eq. 4), $n-1$ parameters for overlapping glottal flows would be required.

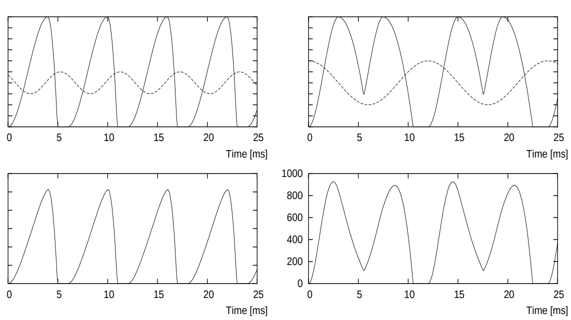

Figure 7: Synthesized laryngeal flows (Left: Drone. Right: Kargyraa). Top: Illustration of the glottal flow by the R-model and modulation. Bottom: Laryngeal flow obtained by modulating glottal flow.

3.2 VTF-modulation+LVT-resonance model

A block diagram of the VTF-modulation + LVT-resonance model is shown in Fig. 8. The model was obtained as follows: The glottal airflow is convoluted with the time-variant laryngeal ventricle resonator depending on the ventricular fold vibration and modulated by the vibration of the ventricular folds.

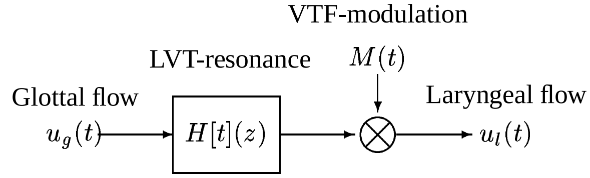

Figure 8: Block diagram of VTF-modulation+LVT-resonance model.

We denote the resonator by the laryngeal ventricle by $H[t](z)$. Then, the laryngeal voice with the laryngeal ventricle resonance (LVT-resonance) and VTF-modulation is described as

$$u_l(t) := M(t) H[t](u_g(t)) \tag{8}$$

We introduce $H[t](z)$ as a time-varying one-pole filter. We calculate the resonance frequency of the laryngeal ventricle, i.e. the frequency of the pole of $H[t]$, based on a Helmholtz resonator. Let $F_v(t)$ be the resonance frequency, d' the thickness of the ventricular folds, and V_{lvt} the volume of the laryngeal ventricle. Then,

$$F_{lvt}(t) = \frac{c}{2\pi} \sqrt{\frac{A'_g(t)}{d' V_{lvt}}} \tag{9}$$

where c is the sound velocity, 3.53×10^4 cm/s.

To permit control flexibility, we define the bandwidth of the resonance by the multiple of variable K, which changes depending on phonation types, and the bandwidth as a Helmholtz resonator. K is used to adjust amplitude of the laryngeal ventricle resonance.

$$B_{lvt}(t) = K \left(\frac{R_{lvt}(t)}{2\pi L_{lvt}(t)} + \frac{G}{2\pi C} \right) \tag{10}$$

The resistance $R_{lvt}(t)$, inductance $L_{lvt}(t)$, conductance G, and capacitance C satisfy the following equations:

$$R_{lvt}(t) = d' \frac{2\sqrt{\pi A'_g(t)}}{A'_g(t)^2} \sqrt{\frac{1}{2}\rho\mu\omega}, \quad L_{lvt}(t) = \frac{\rho d'}{A'_g(t)}$$

$$G = \frac{2 d_{lvt} \sqrt{\pi V_{lvt}} (\eta - 1)}{\rho c^2} \sqrt{\frac{\lambda \omega}{2 \xi \rho}}, \quad C = \frac{V_{lvt}}{\rho c^2}$$

where $\omega := 2\pi/T_0$ is the frequency of the vocal fold vibration, and d_{lvt} the thickness of the laryngeal ventricle. The constants are set as follows: density of air $\rho = 1.14 \times 10^{-3}$ g/cm^3; viscosity $\mu = 1.86 \times 10^{-4}$ dyn $\cdot$ s/cm^2; adiabatic gas constant $\eta = 1.4$; and specific heat $\xi = 0.24$ cal/gm $\cdot$ degree.

4 Model fitting

The proposed model of laryngeal flow with ventricular resonance and supraglottal vibration has many parameters. Moreover, the same laryngeal flow can be obtained from different parameter settings, i.e., a parameter setting and laryngeal flow do not have one-to-one correspondence. Therefore, there are many possibilities to obtain optimal approximation by using this model for the inverse-filtered laryngeal flow. Here, we propose simple fitting methods for laryngeal voices of drone and kargyraa. In the fitting procedure for drone and kargyraa, the VTF-modulation model is used. Once parameters of the VTF-modulation model are fitted, the LVT-resonance model is employed and the resonance of the laryngeal ventricle to the VTF-model are added.

4.1 Model fitting for drone

The fitting of model parameters for the drone voice or other non-subharmonic voices is manually processed. The process consists of the following steps:

Step 1. Estimate F_0 and determine T_0.

Step 2. Adjust O_qT_0 to the first dip after the closing phase.

Step 3. Adjust S_q to match the peak of the model flow to the peak of the flow obtained by inverse-filtering.

Step 4. Adjust ${A_g}_0', \alpha', \theta'$ to minimize the root-mean square error between the flow obtained by inverse-filtering and that obtained using model.

In Fig. 9, the solid line and dotted line show the laryngeal flow obtained with the model and that obtained by inverse-filtering, respectively. For the left panel, the estimated parameters are $O_q = 0.77$, the speed quotient $S_q := t_p/t_n = 1.5$, ${S_q}' = 1.5$, $\theta' = 6\pi/25$, $A'_{g_0} = 0.06$ cm, and $\alpha' = 0.04$ cm. For the right panel, they are $O_q = 0.72$, $S_q = 2.57$, ${S_q}' = 1$, $\theta' = \pi/25$, $A'_{g_0} = 0.06$ cm, and $\alpha' = 0.04$ cm.

For the calculation of the LVT-resonance in both flows, we used $d' = 1$ cm, $d_{lvt} = 0.5$ cm, and $V_{lvt} = 0.5$ cm^3. Parameter K for adjusting bandwidth of the LVT-resonance was set to 0.2.

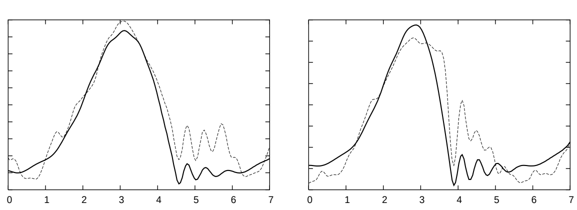

Figure 9: Fitting for drone voices of two different singers. Solid line: Laryngeal flow obtained by the model. Dotted line: Larngeal flow obtained by inverse-filtering.

4.2 Model fitting for kargyraa

Here, only double-periodic subharmonic laryngeal flows are treated. The process of fitting the model parameters for the kargyraa voice or other subharmonic voices consists of the following steps:

Step 1. Estimate F_0 and determine T_0.

Step 2. Determine an interval of the vocal fold closure $[2T_0 - O_qT_0(T_q + 1),\ 2T_0]$.

Step 3. Set $O_q = 1$ and calculate T_q.

Step 4. Adjust $\theta', S_q, {S_q}'$ to approximate the temporal positions of the dip and peak of the model to the flow obtained by inverse-filtering.

Step 5. Adjust $\alpha', {A_g}_0'$ to fit the shape of the flow by the model to the flow obtained by inverse-filtering.

In Fig. 10, the solid line and dotted line show the laryngeal flow by model and laryngeal flow obtained by inverse-filtering, respectively. In the left of Fig. 10, the estimated parameters are as follows: $S_q = 1.22$, ${S_q}' = 1.22$, $\theta' = 9\pi/5$, $A'_{g_0} = 0.12$ cm, $\alpha' = 0.08$ cm. In the right of Fig. 10, the estimated parameters are as follows: $S_q = 1.44$, ${S_q}' = 1.22$, $\theta' = 9\pi/5$, $A'_{g_0} = 0.19$ cm, and $\alpha' = 0.08$ cm.

For calculation of the LVT-resonance in both flows, we use the following parameters: $d' = 1$ cm, $d_{lvt} = 0.5$ cm, and $V_{lvt} = 0.75$ cm^3. A parameter K for adjusting bandwidth of the LVT-resonance is set to 0.2.

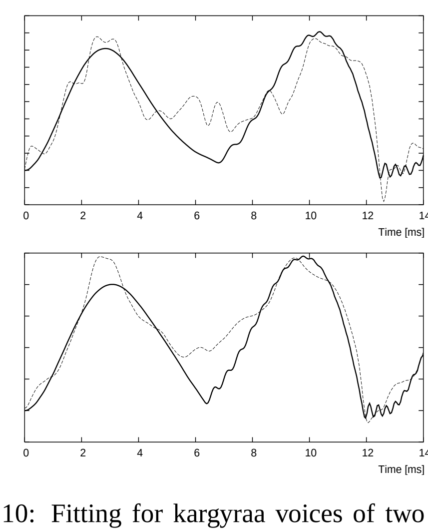

Figure 10: Fitting for kargyraa voices of two different singers. Solid line: Laryngeal flow by the model. Dotted line: the flow obtained by inverse-filtering.

5 Discussions and conclusions

A many-parameter model of the laryngeal flow was proposed. The model consists of a glottal flow model, laryngeal ventricle resonator, and supraglottal modulator. The results show that the proposed model provides a good fit to the estimated laryngeal flow in vocal-ventricular phonations, such as drone and kargyraa. In this paper, we did not mention fitting the model to other pressed-type phonations, such as growl and harsh. However, by changing the LVT-resonance parameters, such as d', d_{lvt}, and V_{lvt}, it would be possible to manipulate the whole epilarynx resonance for vocal-aryepiglottic phonation.

At present, the fitting process for the model is partially done by hand. An automatic fit-estimation procedure of the model will be addressed as future work. Especially, based on perceptual evaluation of importance of the parameters, such as studied in [9], a perceptually effective parameter estimation method is required, because there are many control parameters in the model.

Acknowledgments

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

References

[1] D. G. Childers and C. Ahn, Modeling the glottal volume-velocity waveform for three voice types, *J. Acoust. Soc. Am.*, 97(1): 505–519, 1995.

[2] D. G. Childers and C. K. Lee, Vocal quality factors: analysis, synthesis, and perception, *J. Acoust. Soc. Am.*, 90(5): 2394–2410, 1991.

[3] J. H. Esling, Pharyngeal consonants and the aryepiglottic sphincter, *J. International Phonetics Association*, 26(2), 65–88, 1996.

[4] G. Fant, J. Liljencrants, and Q.-g. Lin, A four-parameter model of glottal flow, *KTH STL-QPSR*, 4/1985:1–14, 1985.

[5] 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.

[6] N. Kobayashi, Y. Tohkura, S. Tenpaku, and S. Niimi, Acoustic and physiological characteristics of traditional singing in Japan, *Tech. Rep. IECE*, SP89-147:39–45, 1990.

[7] P.-Å. Lindestad, M. Sodersten, B. Merker, and S. Granqvist, Voice Source Characteristics in Mongolian "Throat Singing" Studied with High-Speed Imaging Technique, Acoustic Spectra, and Inverse Filtering, *J. Voice*, 15(1): 78–85, 2001.

[8] H.-L. Lu, *Toward a high-quality singing synthesizer with vocal texture control*, Ph D thesis, CCRMA, Stanford Univ., 2002.

[9] P. Mokhtari, Parameterisation and control of laryngeal voice quality by principal components of glottal waveforms, *J. Phonetic Soc. Jpn.*, 7(3):40–54, 2003.

[10] A. E. Rosenberg, Glottal shape on the quality of natural vowels, *J. Acoust. Soc. Am.*, 49(2):583–590, 1970.

[11] K.-I. Sakakibara, L. Fuks, H. Imagawa, and N. Tayama, Growl voice in pop and ethnic styles, *Proc. International Symposium on Musical Acoustics 2004*, 2004.

[12] 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. International Computer Music Conf. 2002*, pp.5–8, 2002.

[13] K.-I. Sakakibara, H. Imagawa, S. Niimi, and N. Osaka, The laryngeal flow model for pressed-type singing voices, *Proc. Stockholm Music Acoustics Conf. 2003*, II:495–498, 2003.

[14] 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 Conf. 2001*, pp. 135–138, 2001.

[15] R. Veldhuis, A computationally efficient alternative for the Liljencrants-Fant model and its perceptual evaluation, *J. Acoust. Soc. Am.*, 103(1):566–571, 1998.