THE LARYNGEAL FLOW MODEL FOR PRESSED-TYPE SINGING VOICES

Ken-Ichi Sakakibara*1, Hiroshi Imagawa*2, Seiji Niimi*3, Naotoshi Osaka*1

*1 NTT Communication Science Laboratories
3-1, Morinosato Wakamiya, Atsugi-shi, Kanagawa, 243-0198, Japan
kis@brl.ntt.co.jp

*2 Department of Speech Physiology, The University of Tokyo
7-3-1, Hongou, Bunkyo-ku, Tokyo, 113-0033, Japan
imagawa@m.u-tokyo.ac.jp

*3 Speech and Hearing Center, International University of Health and Welfare
2600-6, Kitakanemaru, Ohtawara, 324-0011, Japan
niimi@iuhw.ac.jp

ABSTRACT

Asian traditional pressed-type singing voices are different from the European traditional singing voice in their timbre and voice production mechanism. In throat singing, the ventricular folds and true vocal folds vibrate, resulting in the generation of the special laryngeal voice. On the other hand, in some other pressed-type singing voices, such as Japanese Min-yoh, the ventricular folds only approximate but do not vibrate.

We propose a new laryngeal flow model incorporating the effect of the ventricular fold vibration and laryngeal ventricle resonance. The model is a combination of the known glottal airflow model (R-model), the laryngeal ventricle resonance (Helmholtz resonator), and the modulation of ventricular fold vibration. We will also demonstrate the relation between model parameters and voice quality. The results show that the proposed model is effective for synthesizing the pressed-type singing voices.

1. INTRODUCTION

Non-interactive parametric glottal models assume that there are no interactions between the glottal source and vocal tract [2]. The glottal source is described by using mathematical equations. Such models are very effective in speech synthesis and coding and therefore have been used in many studies. The R-model [8] and LF-model [3] have become reference models for this type of model.

All of these models assume that the laryngeal voice source is determined by vocal fold vibratory patterns and intend to control voice quality by changing vocal fold vibratory parameters, such as the open quotient (OQ), speed quotient (SQ), closing quotient (CQ), and amplitude quotient (AQ) [1].

However, in throat singing, the vibration of the ventricular folds (VTFs) (also referred to as the false vocal fold) and strong constriction of the supraglottic structure are observed [11], and in some Asian traditional pressed-type singing, such as Japanese Min-yoh, the constriction of the supraglottic structure is also observed, though the VTFs do not vibrate [5]. Therefore, for synthesis of various styles of singing voices, besides the vocal fold vibration, the effects of the VTF vibration and resonance of the laryngeal ventricle must be considered.

In this paper, we propose a new laryngeal model based on glottal flow, laryngeal ventricle resonance, and the modulation of the VTF vibration. As the laryngeal source for the source-filter synthesis, the proposed model is able to control various timbres of singing voices.

2. THE LARYNGEAL FLOW MODEL WITH VENTRICULAR-FOLD VIBRATORY MODULATION

2.1. VTF-modulation model

VTF vibration is observed in various types of phonation. In throat singing, both drone and kargyraa voice phonations are always accompanied by VTF vibrations, as well as vocal fold (VF) vibrations. In the drone voice, the ventricular folds vibrate in the same period as the VFs, and in the kargyraa voice, the VTF vibrate in an integer multiple (usually double or triple) period of the VFs [4, 6, 11]. The results of a simulation using a 2×2-mass model suggest the possible vibratory patterns of the VFs and VTFs [9, 11].

Here, we use “laryngeal flow (source)” to mean the airflow through the VTF slit, and “glottal airflow (source)” to mean the airflow through the slit of the VFs.

The laryngeal flows of drone and kargyraa for different two singers are shown in Fig. 1. These flows were obtained from recorded sounds using an inverse-filter analysis. We marked five poles on spectrum in the range from 0 to 5 kHz, constructed the inverse-filter, and manually adjusted it to make the result smooth. By combining the results of high-speed images, EGG waveforms, and these inverse-filtered laryngeal sources, we concluded that, in throat singing, the VTF vibration is indispensable for the generation of the laryngeal flow. Therefore, modeling the laryngeal flow in throat singing requires a new laryngeal model that includes the effect of VTF vibration.

The VTF-modulation model \( \tilde{u}(t) \) is simply defined as follows:

\[
\tilde{u}(t) := M(t)u(t)
\]  \( (1) \)

The block-diagram of the model is depicted as shown in Fig. 2.

In this paper, we choose a simple R-model [8] for the glottal flow. The R-model is described as follows:

\[
u(t) = \begin{cases} 
\alpha \left( \frac{3}{r_p} \right)^2 - 2 \left( \frac{1}{r_p} \right)^3, & \text{if } 0 \leq t \leq T_p \\
\alpha \left( 1 - \left( \frac{t - T_p}{r_p} \right)^2 \right), & \text{if } T_p \leq t \leq T_p + T_o \leq T_0
\end{cases}
\]  \( (2) \)
appropriate model with laryngeal ventricle resonance is required. Fig. 3 with the physical model simulation results, [7], hence, some appro-

Figure 1: Inverse-filtered laryngeal flow. Left upper: drone by subject A, left lower: drone by subject B, right upper: karygyraa by A, right lower karygyraa by B.

VTF-modulation

\[
\begin{align*}
M(t) & \quad \text{Glottal flow} \\
\quad u(t) & \quad \text{Laryngeal flow} \\
& \quad u'(t)
\end{align*}
\]

Figure 2: Block diagram of VTF-modulation model

where \(\alpha\) is amplitude, \(T_p\) opening time, \(T_n\) closing time, and \(T_0\) period. All of these variables are in \(\mathbb{R}_{>0}\). The open quotient (OQ) is written as \((T_p + T_n)/T_0\).

The vibratory patterns of the VTFs were observed using the high-speed images and seem to be not exactly sine-shape [7, 10, 11]. However, here we define the VTF-modulation function \(M(t)\) by multiplication by constant \(M\) of the false glottal area function \(A'_{y_{0}}\). We also define \(A'_{y_{0}}\) as a sine function:

\[
\begin{align*}
M(t) & := M A'_{y_{0}}(t) \\
A'_{y_{0}}(t) & := \alpha' \sin(\omega' t + \theta') + A'_{y_{0}}
\end{align*}
\]

(3)

where \(\alpha'\) represents the amplitude of the VTF vibration, \(A'_{y_{0}}\) the area between the VTFs at rest, \(\omega'\) the frequency of the VTF vibration, and \(\theta'\) the phase difference of the VTF vibration from VF vibration. All of these are in \(\mathbb{R}_{>0}\). Physiological observations and the simulation using 2x2 mass model suggest that the periods of the VF and VTF vibration satisfy \(2\pi/\omega' = \pi T_0\), where \(n \in \mathbb{Z}_{>0}\).

2.2. VTF-modulation and LVT-resonance model

The laryngeal ventricle is the space between the VFs and VTFs. When the VTFs are strongly constricted, it seems the effect of this small space on the laryngeal voice can not be ignored. The physical model simulation suggest that some acoustic effects occur around 2000 Hz [9, 11]. The inverse-filtered laryngeal voices of throat singing have some ripples (Fig. 1), which almost agree with the physical model simulation results, [7], hence, some appropriate model with laryngeal ventricle resonance is required. Fig. 3 shows spectra of the drone voices of two different singers.

A block diagram of our proposed model (VTF-modulation and LVT-resonance model) is shown in Fig. 4.

The model was obtained as follows: The glottal airflow is convoluted with the time-variant laryngeal ventricle resonator depending on the VTF vibration, and modulated by the vibration of the VTFs.

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

\[
\dot{u}(t) := M(t) H[t](u(t))
\]

(5)

We realize \(H[t](z)\) as a time-varying one-pole filter. We calculate the resonance frequency of the laryngeal ventricle, i.e. the frequency of the pole of \(H[t]\), by means of a Helmoltz resonator. Let \(F_r(t)\) be the resonance frequency, \(d\) be thickness of the VTF, and \(V_v\) be volume of the laryngeal ventricle. Then,

\[
F_r(t) = \frac{c}{2\pi} \sqrt{\frac{A_0(t)}{d V_v}}
\]

(6)

where \(c\) is the sound velocity, \(3.53 \times 10^4\) 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 Helmoltz resonator.

\[
B_r(t) = \frac{R_r(t)}{2\pi L_r(t)} + \frac{G}{2\pi C}
\]

(7)

The resistance \(R_r(t)\), inductance \(L_r(t)\), conductance \(G\), and capacitance \(C\) satisfy the following equations.

\[
R_r(t) = d R_{0r} \frac{2\sqrt{\pi A_0(t)}}{A_0'(t)} \sqrt{\frac{1}{\pi \mu \omega}}, \quad L_r(t) = \frac{\rho f}{A_0'(t)}
\]

\[
G = \frac{2d_v \sqrt{\pi V_v(\eta - 1)}}{\rho c^2} \sqrt{\frac{\lambda \omega}{2\xi \rho}}, \quad C = \frac{V_v}{\rho c^2}
\]

(8)

where \(\omega := 2\pi/T_0\) is the frequency of the VF vibration, and \(d_v\) the thickness of the laryngeal ventricle. The constants are set as follows: the density of air \(\rho = 1.14 \times 10^{-3}\) g/cm\(^3\); the viscosity \(\mu = 1.86 \times 10^{-5}\) dyn · s/cm\(^2\); the adiabatic gas constant \(\eta = 1.4\); and the specific heat \(\xi = 0.24\) cal/gm · degree.
3. ACOUSTICAL CHARACTERISTICS

3.1. VTF-modulation model

We study the effect of the phase difference between the VF and VTF vibrations. In the equation

\[ u'(t) = (\alpha' \sin(\omega t + \theta') + A'_{\theta_0}) u(t), \tag{8} \]

we fix \( A'_{\theta_0} = 0.5 \max u(t) \), \( \alpha' = 0.35 \max u(t) \). For \( u(t) \), we also set \( T_0 = 8 \text{ ms}, T_p/T_0 = 0.42 \), \( T_c/T_0 = 0.18 \), and hence \( \text{OQ} = 0.6 \). As these settings, the spectral tilt of \( u(t) \) is close to 12 dB. We also set \( \omega' = \omega \), i.e. study the laryngeal flow, such as the drone voice of throat singing. The laryngeal flows of various \( \theta' \) are as shown in Fig. 5.

![Figure 5: Laryngeal flows in various phases.](image)

In Fig. 5, the laryngeal flow for \( \theta' = \pi/4 \) is also supported by its frequent appearance in physical model simulation [9]. In Fig. 5, the laryngeal flow for \( \theta' = \pi/4 \) shows the similar characteristics of the drone in Fig. 1 and the opening duration is relatively less than the closing duration.

Fig. 6 shows the spectral envelops of the laryngeal flows for different \( \theta' \). Among \( \theta' \)'s in Fig. 6, the spectral tilt is the largest when \( \theta' = -\pi/4 \) (-16 dB/octave), smallest when \( \theta' = \pi/2 \) (-12 dB/octave).

3.2. VTF-modulation and LVT-resonance model

We set \( A'_{\theta_0} = 0.10 \text{ cm} \) and \( \alpha' = 0.05 \text{ cm} \). We normalize \( u(t) \) by multiplying some real positive value and assume \( u(t) \) as the glottal area function. We set the maximal glottal area \( \max u(t) \) to 0.2 cm\(^2\). We set the thickness of the VTF \( d' \) to 1.0 cm, the cross-sectional area of the laryngeal ventricle to 1.5 cm\(^2\), and the depth of the laryngeal ventricle to 0.5 cm, \( K = 20 \) in Eq. (7). We used these values for calculation of the \( H[f](\omega) \). The other values are the same as above. The laryngeal flows with LVT resonance of various \( \theta' \)'s are as shown in Fig. 7.

In all cases, ripples are observed after the closure of the glottis. Fig. 8 shows spectra of two flows. The effect of the VTF resonance is observed around 2000 Hz. This feature is observed in all the synthesized sources.

![Figure 7: Laryngeal flows in various phases, with laryngeal ventricle resonance.](image)

![Figure 8: Spectra of the synthesized laryngeal voice.](image)

3.3. False glottal area at rest

When \( A'_{\theta_0} \) is decreased, Eq. (6) implies that the resonance frequency is pushed higher. The spectra in Fig. 9 shows the spectra for different \( A'_{\theta_0} \). Other conditions are the same as above.
behavior similar to kargyraa phonation. From the characteristics vibrations we synthesized laryngeal flows by changing the amplitude of VTF modulations.

3.4. Modulation amplitude

We synthesized laryngeal flows by changing the amplitude of VTF vibrations $\alpha'$. No significant trends are observed in the behaviors of the synthesized flows.

3.5. Laryngeal source for kargyraa

If $\omega' = 2\omega$, then $u'(t)$ has a double-period of $u(t)$ and shows behavior similar to kargyraa phonation. From the characteristics of $u(t)$, in the middle of each period, the laryngeal flow reach to 0. However, the inverse-filtered kargyraa voice maintains flow in each period. Uncompleted closure of the VFAs are also observed in the physical model simulation [9]. In order to obtain the similar laryngeal flow shape, the second $u(t)$ flow must start before $T_o + T_p$ or $u(t)$ needs sufficiently large OQ.

4. CONCLUSIONS

A new laryngeal flow model was proposed. We studied the acoustic characteristics of the model by changing its parameters. To obtain the laryngeal voice shape of the drone voice, VTF modulation is indispensable. In addition, to obtain ripples after the closure of the vocal folds, laryngeal ventricle resonance is effective. These results show the proposed model is effective for synthesizing pressed-type singing voices, such as throat singing. Parameter fitting in terms of analysis-by-synthesis and perceptual evaluation will be addressed as future works. In addition, an effective inverse-filtering method in cases that the source has poles and the filter has zeros must be studies.

Acknowledgments

We thank Seiji Adachi, Parham Mokhtari, Yoshinao Shiraki, Niro Tayama, and Masahiko Todoriki for their helpful discussions.

5. REFERENCES