

Voice production model integrating boundary-layer analysis of glottal flow and source-filter coupling

Kaburagi, Tokihiko

Faculty of Design, Department of Communication Design Science, Kyushu University

<https://hdl.handle.net/2324/25534>

出版情報 : Journal of the Acoustical Society of America. 129 (3), pp.1554-1567, 2011-03.
Acoustical Society of America

バージョン :

権利関係 : (C) 2011 Acoustical Society of America.



Voice production model integrating boundary-layer analysis of glottal flow and source-filter coupling

Tokihiko Kaburagi^{a)}

Faculty of Design, Department of Communication Design Science, Kyushu University, 4-9-1 Shiobaru, Minami-ku, Fukuoka 815-8540, Japan

(Received 11 April 2010; revised 20 September 2010; accepted 12 December 2010)

A voice production model is created in this work by considering essential aerodynamic and acoustic phenomena in human voice production. A precise flow analysis is performed based on a boundary-layer approximation and the viscous–inviscid interaction between the boundary layer and the core flow. This flow analysis can supply information on the separation point of the glottal flow and the thickness of the boundary layer, both of which strongly depend on the glottal configuration and yield an effective prediction of the flow behavior. When the flow analysis is combined with the modified two-mass model of the vocal fold [Pelorson *et al.* (1994). *J. Acoust. Soc. Am.* **96**, 3416–3431], the resulting acoustic wave travels through the vocal tract and a pressure change develops in the vicinity of the glottis. This change can affect the glottal flow and the motion of the vocal folds, causing source-filter coupling. The property of the acoustic feedback is explicitly expressed in the frequency domain by using an acoustic tube model, allowing a clear interpretation of the coupling. Numerical experiments show that the vocal-tract input impedance and frequency responses representing the source-filter coupling have dominant peaks corresponding to the fourth and fifth formants. Results of time-domain simulations also suggest the importance of these high-frequency peaks in voice production.

© 2011 Acoustical Society of America. [DOI: 10.1121/1.3533732]

PACS number(s): 43.70.Bk, 43.70.Gr [MAH]

Pages: 1554–1567

I. INTRODUCTION

The process of human voice production is highly complex because of the interactions among the underlying physical phenomena including the volumetric flow that passes through the glottis, the mechanical motion of the vocal fold, and the propagation of acoustic waves in the vocal tract. The original two-mass model (Ishizaka and Flanagan, 1972) has allowed great progress into understanding the process of the voice production, particularly by using the separation-reattachment scheme of flow in a confined channel with a constriction (Broad, 1979). Because the temporal change of glottal flow serves as the pressure source of speech, the dynamics of viscous flow in the glottis have been studied extensively using flow measurements (van den Berg *et al.*, 1957; Scherer *et al.*, 1983; Hofmans *et al.*, 2003), computer simulations (Ishizaka and Flanagan, 1972; Broad, 1979; Pelorson *et al.*, 1994; Story and Titze, 1995; Lous *et al.*, 1998), and numerical solutions of the conservation equations (Zhao *et al.*, 2002).

Although the Bernoulli relation is capable of predicting the flow in the glottis (van den Berg *et al.*, 1957), the air viscosity can not be ignored completely. This is because a boundary layer develops along the surface of the vocal folds and detaches from the glottal wall at the abrupt expansion of the glottal exit (Ishizaka and Flanagan, 1972). It is known that convergent–divergent changes in the glottal shape control the aerodynamic pressure inside the glottis and

transform the flow energy into vocal-fold oscillation efficiently by moving the flow separation point (Titze, 1988; Zhang *et al.*, 2006). Such a movable nature of the separation point was successfully introduced using a boundary-layer theory (Pelorson *et al.*, 1994; Lous *et al.*, 1998). In the effective one-mass expression (Zañartu *et al.*, 2007), the effect of the flow separation and influence of the boundary-layer thickness were implicitly taken into consideration in terms of the orifice discharge coefficient.

To overcome the problem of the movable separation point and consider the influence of the boundary-layer thickness, we numerically solved the Kármán equation on the basis of the Pohlhausen method (Schlichting and Gersten, 2000) for one- and two-dimensional glottal channels (Kaburagi, 2008; Kaburagi and Tanabe, 2009). The effective flow channel is then closely estimated by subtracting the displacement thickness of the boundary layer from the nominal channel space bounded by the vocal folds. Under a static pressure difference between the sub- and supra-glottal portions, the resulting volume flow is closely related to the effective glottal area at the separation point. Furthermore, the Kármán–Pohlhausen framework gives a proper solution for general configurations of a smoothly shaped glottal channel. A mass-spring model with a rounded configuration (Pelorson *et al.*, 1994), a distributed-mass model (Titze, 1973), or a continuum model (Berry and Titze, 1996) can then be used to represent the mechanical motion of the vocal folds in combination with the boundary-layer approximation. The flow and vocal folds can interact mutually through flow-driven mechanical oscillations and shaping of the glottal channel.

^{a)}Author to whom correspondence should be addressed. Electronic mail: kabu@design.kyushu-u.ac.jp

In addition, the source-filter coupling generates nonlinear effects such as flow skewing (Rothenberg, 1981), enhancement or suppression of vocal-fold oscillation (Zañartu *et al.*, 2007; Titze, 2008), lowering of the oscillation threshold pressure (Chan and Titze, 2006), discontinuity of the fundamental frequency change (Ishizaka and Flanagan, 1972; Švec *et al.*, 1999), and generation of bifurcation and voice instabilities (Tokuda *et al.*, 2007). Time-domain computer simulation has been a major tool in investigating these coupling effects, and the consequences have been analyzed in the frequency domain as an interrelation between the fundamental frequency of the voice and the resonant and antiresonant frequencies of the vocal tract (Ishizaka and Flanagan, 1972). Titze (2008) and Titze and Worley (2009) interpreted them more specifically in terms of the reactance of impedances arising from the acoustic load of the sub- and supra-glottal tracts.

This paper presents a voice production model that can explain essential aerodynamic and acoustic phenomena in phonation. Together with the interactive boundary-layer analysis (Kaburagi, 2008), an acoustic tube model (Sondhi and Schroeter, 1987) is employed for the sub- and supra-glottal tracts and a mass-spring model (Pelorson *et al.*, 1994) for the vocal folds. Following the theory of Titze (2008), we connected the acoustic tubes of the sub- and supra-glottal tracts to the intermediate part of the glottis so as to examine the frequency response of the acoustic pressures near the glottis relative to each acoustic component of the glottal volume flow. With this acoustic impedance representation, the effects of the acoustic pressure difference across the glottis and the mean intra-glottal pressure are interpreted. The phase components of these frequency responses provide information about the relative timing between the flow and pressure values, as does a reactance curve analysis (Titze, 2008; Titze and Worley, 2009). The phase responses therefore determine the quality of the source-filter coupling. The magnitude responses, on the other hand, give an indication of the quantity and strength of the coupling, since they represent the amplitudes of the pressure values fed back from the tracts for unit-amplitude frequency components of the glottal flow.

In our voice production model, these frequency characteristics are directly integrated with the flow analysis and vocal-fold model by taking their inverse Fourier transform and convolving the result with the glottal flow waveforms. Mathematical expressions are then derived to estimate the glottal volume flow and the driving force of the vocal folds from the aerodynamic and acoustic pressures. Numerical investigations are performed using the area function of the vocal tract (Story *et al.*, 1996), with a special emphasis on the frequency characteristics of the feedback acoustic pressures. Detailed interpretations of these characteristics are given and they are used to explain the influences of the source-filter coupling. The voice production model is then used for time-domain simulations, and the influence of the source-filter coupling is examined by changing the frequency characteristics of the feedback acoustic pressures.

This paper is organized as follows. Section II provides mathematical explanations of the voice production model,

including the flow analysis method, the vocal-fold model, the acoustic model of the sub- and supra-glottal tracts, and the interrelations among the models. The simulation procedure is also given. Numerical results are shown in Sec. III. Finally, the conclusions of this work are given in Sec. IV.

II. VOICE PRODUCTION MODEL INTEGRATING BOUNDARY-LAYER ANALYSIS AND SOURCE-FILTER COUPLING

This section presents a mathematical explanation of our voice production model. The model is composed of a method of estimating glottal volume flow, a mechanical model of the vocal folds, and an acoustic tube model of the sub- and supra-glottal tracts. Given the pressure difference between the sub- and supra-glottal regions, the volume flow passing through the glottis is estimated by a boundary-layer approximation that includes the interaction between the core flow and the boundary layer (Kaburagi, 2008; Kaburagi and Tanabe, 2009). Here, the pressure difference across the glottis is determined by summing the static sub-glottal pressure supplied by the lungs and the acoustic pressure representing the source-filter coupling. The pressure distribution along the glottal channel is predicted using the flow analysis, allowing the driving force of the vocal folds to be determined. In addition, the acoustic pressure near the glottis drives the vocal folds. Thus, another type of source-filter coupling is considered. Movement of the vocal folds is modeled by the modified two-mass model (Pelorson *et al.*, 1994). The acoustic transmission property in each tract is modeled in the frequency domain using the acoustic tube model of Sondhi and Schroeter (1987).

A. Glottal volume flow estimation

A method for estimating the glottal volume flow that incorporates the influence of air viscosity and acoustical coupling on the sub- and supra-glottal tracts is proposed below. Since the Reynolds number of glottal flow is typically of the order of 10^3 , the boundary-layer approximation (Schlichting and Gersten, 2000) is applicable as the basic framework for analyzing glottal flow together with flow separation. From previous studies (Pelorson *et al.*, 1994; Kaburagi and Tanabe, 2009), quasi-steady, incompressible, and one-dimensional assumptions can be made. In addition, we assume that movements of the vocal folds, and hence the glottal channel, are symmetrical with respect to the x axis, as illustrated in Fig. 1.

Given the pressure difference, Δp , between the inlet (x_0) and outlet (x_e) of the glottis and the cross-sectional area of the glottal outlet, S , glottal volume flow, u_g , may be estimated based on the Bernoulli relation, $\Delta p = 0.5\rho u_g^2/S^2$, where ρ is the air density. However, flow analyses based on the boundary-layer approximation have shown that the pressure difference is more precisely related to the cross-sectional area at the flow separation point (Pelorson *et al.*, 1994). In addition, the boundary layer can reduce the effective size of the glottal channel. Therefore, the displacement thickness of the boundary layer can be excluded from the

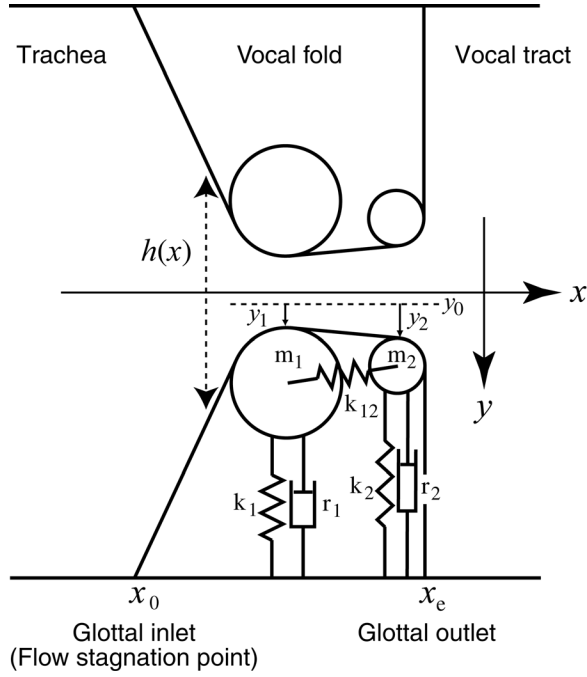


FIG. 1. Representation of the one-dimensional glottal channel and the mechanical model of the vocal folds. The channel and the vocal-fold motion are symmetrical with respect to the x axis. $h(x)$ is the channel height. The y axis perpendicular to x is used to represent the mass displacements. The length of the vocal fold, which is perpendicular to the $x - y$ plane, is l_g .

nominal configuration of the channel. From these considerations, we obtain a refined pressure-flow relation

$$\Delta p = 0.5\rho \frac{u_g^2}{S_s^2}. \quad (1)$$

S_s is the effective channel area at the separation point and is given by

$$S_s = l_g \{h(x_s) - 2\delta(x_s)\}, \quad (2)$$

where l_g is the length of the vocal fold, $h(x)$ is the channel height, $\delta(x)$ is the displacement thickness, and x_s is the point of flow separation along the x axis. The glottal area and other variables in Eq. (1) usually vary in time, but based on the quasi-steady assumption, they are treated as time-independent variables at each simulation time step.

The pressure difference generally includes both aerodynamic and acoustic components, Δp_F and Δp_A , as

$$\Delta p = \Delta p_F + \Delta p_A. \quad (3)$$

The aerodynamic pressure becomes small enough downstream of the separation point, so that $\Delta p_F \approx p_{F0}$, where p_{F0} is the sub-glottal pressure. Δp_A is the acoustic pressure difference across the glottis, a pressure gradient that induces flow. This is explained in Sec. II C but is given by a convolution form

$$\Delta p_A = z_\Delta(t) * u_g(t)|_{t=t_c} = \int z_\Delta(\tau) u_g(t_c - \tau) d\tau, \quad (4)$$

where $z_\Delta(t)$ is the impulse response of a linear, causal filter that predicts the pressure difference from the waveform of the glottal volume flow. Note that the convolution is performed for the current (latest) time instant, $t = t_c$. As discussed later, $z_\Delta(t)$ is derived by applying the lossy acoustic tube model to the sub- and supra-glottal tracts and the glottis. The acoustic properties of this filter depend on the cross-sectional area of these cavities, but their time dependencies, due to the deformation of the vocal tract or the glottis, are implicit. We therefore suppose that $z_\Delta(t)$ is determined by the area function of the cavities measured at the time t_c .

By combining Eqs. (1) and (3), and the expressions for Δp_F and Δp_A , we obtain the following relation:

$$p_{F0} + z_\Delta(t) * u_g(t)|_{t=t_c} = 0.5\rho \frac{u_g^2}{S_s^2}. \quad (5)$$

It can be rewritten in discrete-time form as

$$p_{F0} + \sum_{k=0}^{K-1} z_\Delta(k) u_g(n-k) = 0.5\rho \frac{u_g(n)^2}{S_s^2}, \quad (6)$$

where K is the length of $z_\Delta(k)$ and n is the time index corresponding to t_c . Therefore, the glottal volume flow can be estimated from Eq. (6) as

$$u_g(n) = \frac{z_\Delta(0)S_s^2 + S_s \sqrt{\{z_\Delta(0)S_s\}^2 + 2\rho P}}{\rho}, \quad (7)$$

where

$$P = p_{F0} + \sum_{k=1}^{K-1} z_\Delta(k) u_g(n-k). \quad (8)$$

These equations indicate that the time history of the glottal flow waveform affects the flow estimation recursively.

The value of S_s in Eq. (7) can be determined from the information on the flow separation point and the displacement thickness of the boundary layer. The interactive boundary-layer method is employed here. The Kármán equation can be expressed as (Kalse et al., 2003)

$$v(x) \frac{\delta(x)}{v} \frac{d}{dx} \frac{\delta(x)}{H(x)} + \left(1 + \frac{2}{H(x)}\right) F_1(H(x)) = H(x) F_2(H(x)) \quad (9)$$

and

$$\frac{\delta(x)^2}{v} \frac{dv(x)}{dx} = F_1(H(x)), \quad (10)$$

where $F_1(H(x)) = -2.4\{1 - \exp(0.43(2.59 - H(x)))\}$ and $F_2(H(x)) = 4/H(x)^2 - 1/H(x)$ are approximation functions. $H(x)$ is the ratio of the boundary-layer thicknesses, given by $H(x) = \delta(x)/\theta(x)$, where $\theta(x)$ is the momentum thickness. v is the kinematic viscosity. The effective core flow velocity can be expressed as

$$v(x) = \frac{u_g}{l_g \{h(x) - 2\delta(x)\}}. \quad (11)$$

Equations (9)–(11) constitute a set of nonlinear simultaneous equations with respect to three unknown variables: $H(x)$, $\delta(x)$, and $v(x)$. It can be solved numerically using a downstream marching technique (Kaburagi, 2008). Using the solution to this problem, the flow separation point, x_s , is estimated by finding the position where the wall shear stress becomes zero, indicating that $F_2 = 0$ and $H = 4$ at that point. The effective cross-sectional area, S_s , is then determined by Eq. (2).

The interactive boundary-layer method takes the channel height, $h(x)$, and the volume flow rate, u_g , as the input variables, where we suppose that l_g is constant. However, the volume flow rate can in turn be estimated by Eq. (7). Because the interactive boundary-layer method and the flow estimation equation are difficult to couple, they are solved alternately, as explained in Sec. II E. After the core flow velocity, $v(x)$, is obtained from the boundary-layer method, the pressure distribution along the glottis, $p_F(x)$, is determined from the Bernoulli relation as

$$p_F(x) = p_{F0} - \frac{1}{2} \rho v(x)^2. \quad (12)$$

B. Mechanical vocal-fold model

Mass-spring models are widely used to represent motion of the vocal folds (Ishizaka and Flanagan, 1972; Broad, 1979; Sondhi and Schroeter, 1987; Pelorson *et al.*, 1994; Story and Titze, 1995; Lous *et al.*, 1998; Zañartu *et al.*, 2007). However, a numerical solution to the Kármán equation requires that the glottis be configured smoothly, to avoid any discontinuities in the flow velocity. Therefore, the smooth model of Pelorson *et al.* (1994) is adopted in the present study. Each vocal fold is constructed by two cylinders, located at the lower and upper parts of the fold, and three plates, which connect the inlet and outlet of the glottis and the cylinders (Fig. 1). Masses m_1 and m_2 are assigned to each cylinder and are connected to the fixed wall by dampers of resistance r_1 and r_2 , and linear springs with Hooke's constants k_1 and k_2 . The two masses are joined by another linear spring of constant k_{12} .

The equation of vocal-fold motion for each mass is given as (Pelorson *et al.*, 1994)

$$m_1 \frac{d^2 y_1}{dt^2} + r_1 \frac{dy_1}{dt} + k_1 y_1 + k_{12}(y_1 - y_2) = f_1 \quad (13)$$

and

$$m_2 \frac{d^2 y_2}{dt^2} + r_2 \frac{dy_2}{dt} + k_2 y_2 + k_{12}(y_2 - y_1) = f_2. \quad (14)$$

y_1 and y_2 are the displacements of the masses perpendicular to the x axis. The absolute mass position is $y_1 + y_0$ and $y_2 + y_0$, where y_0 is the common rest position. The stiffnesses of the springs are linear even when the vocal folds collide with each other. The effect of collision is incorporated by increasing the values of the mechanical constants of the springs and dampers. In accordance with the literature (Pelorson *et al.*, 1994), the driving force from the glottal flow and the acoustic wave

is exerted only on the lower, first mass, and f_2 is set to zero. The driving force, f_1 , is determined as follows.

When $y_1 + y_0 > 0$ and $y_2 + y_0 > 0$ hold, the glottis is open and the distribution of the aerodynamic pressure, $p_F(x)$, is given by Eq. (12). The acoustic driving force is estimated as the mean acoustic pressure between the inlet and outlet of the glottis. Similar to Eq. (4), this mean acoustic pressure is defined as

$$\bar{p}_A = z_M(t) * u_g(t)|_{t=t_c} = \int z_M(\tau) u_g(t_c - \tau) d\tau. \quad (15)$$

$z_M(t)$ is the impulse response of a linear, causal filter determined by the area function of the sub- and supra-glottal tracts and the glottis, as explained in Sec. II C. The entire driving pressure is then given as

$$p(x) = p_F(x) + \bar{p}_A = p_{F0} - \frac{1}{2} \rho v(x)^2 + z_M(t) * u_g(t) \Big|_{t=t_c} \quad (16)$$

and the driving force for the lower mass as

$$f_1 = \lambda l_g \left\{ \int_{x_0}^{x_s} p_F(x) dx + (x_e - x_0) \bar{p}_A \right\}. \quad (17)$$

Here, x_s is the flow separation point and λ is a parameter specifying the area upon which the pressure is exerted. If the glottis is closed, f_1 is set to $f_1 = \lambda l_g (x_c - x_0) p_{F0}$.

C. Frequency-domain interpretation of the source-filter coupling

In this subsection, an acoustic tube model (Sondhi and Schroeter, 1987) is applied to calculate properties of the sub- and supra-glottal tracts. In addition, the difference and mean of the acoustic pressures at the glottal inlet, p_{A0} , and outlet, p_{A1} , are estimated to determine the effects of source-filter coupling. The pressure difference, Δp_A in Eq. (4), is defined as

$$\Delta p_A = p_{A0} - p_{A1} \quad (18)$$

and the mean pressure, \bar{p}_A in Eq. (15), as

$$\bar{p}_A = \frac{p_{A0} + p_{A1}}{2}. \quad (19)$$

To discriminate between these temporal waveforms and their Fourier transforms in the following expressions, time-domain variables are denoted by lower-case letters while frequency-domain variables are denoted by capital letters.

The signal flow graph in Fig. 2 shows the acoustic model used in this study. The glottis configuration is simply modeled by a uniform acoustic tube with length L_g , which corresponds to the glottal depth, and cross-sectional area $S_g = l_g \sqrt{(y_1 + y_0)(y_2 + y_0)}$, which is the geometric average between the lower and upper parts of the glottis. The input and output variables for the glottis are then expressed as

$$\begin{pmatrix} P_{A1} \\ U_{A1} \end{pmatrix} = \begin{pmatrix} A_g & B_g \\ C_g & D_g \end{pmatrix} \begin{pmatrix} P_{A0} \\ U_{A0} \end{pmatrix}, \quad (20)$$

where P_{A0} and P_{A1} are the Fourier transforms of the acoustic pressures and U_{A0} and U_{A1} those of the volume velocities. Note that these physical variables and the components of the transmission matrix are a function of angular frequency, ω . The transmission matrix can be obtained from the acoustic tube model (Sondhi and Schroeter, 1987) as $A_g = \cosh(\sigma L_g/c)$, $B_g = -(\rho c/S_g)\gamma \sinh(\sigma L_g/c)$, $C_g = -(S_g/\rho c)(\sinh(\sigma L_g/c)/\gamma)$, and $D_g = \cosh(\sigma L_g/c)$, where c is the speed of sound. $\alpha = \sqrt{j\omega c_1}$, $\beta = j\omega\omega_0^2/\{(j\omega + a)j\omega + b\} + \alpha$, $\gamma = \sqrt{(\alpha + j\omega)/(\beta + j\omega)}$, and $\sigma = \gamma(\beta + j\omega)$ are frequency-dependent parameters, where $a = 130\pi$ rad/s, $b = (30\pi)^2$ (rad/s)², $c_1 = 4$ rad/s, and $\omega_0^2 = (406\pi)^2$ (rad/s)². In addition, the following relationships are obtained from Fig. 2:

$$P_{A0} = -Z_0 U_{A0}, \quad (21)$$

$$P_{A1} = Z_1 U'_{A1}, \quad (22)$$

and

$$U'_{A1} = U_{A1} + U_g, \quad (23)$$

where Z_0 and Z_1 are the input impedances of the sub- and supra-glottal tracts, respectively.

From Eqs. (20) and (21), the input impedance of the glottis seen from the vocal tract, Z_g , can be expressed as

$$Z_g = -\frac{P_{A1}}{U_{A1}} = -\frac{A_g P_{A0} + B_g U_{A0}}{C_g P_{A0} + D_g U_{A0}} = -\frac{B_g - A_g Z_0}{D_g - C_g Z_0}. \quad (24)$$

The relations

$$U_g = U'_{A1} - U_{A1} = \frac{P_{A1}}{Z_1} + \frac{P_{A1}}{Z_g} = \frac{Z_1 + Z_g}{Z_1 Z_g} P_{A1} \quad (25)$$

and

$$P_{A1} = A_g P_{A0} + B_g U_{A0} = \left(A_g - \frac{B_g}{Z_0}\right) P_{A0} \quad (26)$$

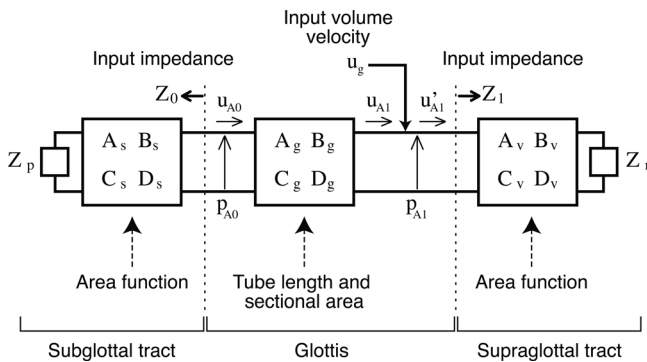


FIG. 2. The acoustic model representing the wave propagation in the sub- and supra-glottal tracts and through the glottis. The transmission matrix of each section is determined by the cross-sectional area and length. Together with the terminal impedances for the lungs and lips, the model allows the calculation of input impedance, the vocal-tract transfer function, and the pressure at both ends of the glottal tube.

are also obtained. Therefore, P_{A0} can be expressed as a function of U_g by eliminating P_{A1} in Eqs. (25) and (26)

$$P_{A0} = \frac{Z_1 Z_g}{Z_1 + Z_g} \cdot \frac{Z_0}{A_g Z_0 - B_g} U_g = \frac{Z_0 Z_1}{Z_D} U_g. \quad (27)$$

P_{A1} can be expressed as

$$P_{A1} = -\frac{(B_g - A_g Z_0) Z_1}{Z_D} U_g, \quad (28)$$

where $Z_D = (D_g - C_g Z_0) Z_1 - (B_g - A_g Z_0)$.

Finally, ΔP_A and \bar{P}_A in Eqs. (18) and (19) are each obtained as the frequency response of a filter having the acoustic component of the glottal volume flow as its input

$$\frac{\Delta P_A}{U_g} = \frac{P_{A0} - P_{A1}}{U_g} = \frac{\{B_g - (A_g - 1)Z_0\}Z_1}{Z_D} \quad (29)$$

and

$$\frac{\bar{P}_A}{U_g} = \frac{1}{2} \cdot \frac{P_{A0} + P_{A1}}{U_g} = -\frac{\{B_g - (A_g + 1)Z_0\}Z_1}{2Z_D}. \quad (30)$$

In the next section, we examine these frequency responses using the area function of the vocal tract (Story et al., 1996) and show the frequency dependencies of the source-filter coupling. $z_A(t)$ in Eq. (4) and $z_M(t)$ in Eq. (15) are the inverse Fourier transforms of $Z_A(\omega) = \Delta P_A(\omega)/U_g(\omega)$ and $Z_M(\omega) = \bar{P}_A(\omega)/U_g(\omega)$, respectively.

The input impedance of the supra-glottal tract, Z_1 , can be obtained, in accordance with Eq. (12) of Sondhi and Schroeter (1987), as

$$Z_1 = \frac{D_v Z_r - B_v}{A_v - C_v Z_r}. \quad (31)$$

If the production of vowel-like sounds is considered, the matrix chaining is

$$\begin{pmatrix} A_v & B_v \\ C_v & D_v \end{pmatrix} = \begin{pmatrix} A_{N_v} & B_{N_v} \\ C_{N_v} & D_{N_v} \end{pmatrix} \begin{pmatrix} A_{N_v-1} & B_{N_v-1} \\ C_{N_v-1} & D_{N_v-1} \end{pmatrix} \cdots \begin{pmatrix} A_1 & B_1 \\ C_1 & D_1 \end{pmatrix}, \quad (32)$$

where the first tube section corresponds to the entry of the tract and the N_v th to the lip end. Components A_i , B_i , C_i , and D_i for the i th tube section are given by the length and cross-sectional area of that section, as for the components of Eq. (20). Z_r is the radiation impedance at the lip opening (Flanagan, 1972).

The sub-glottal tract can also be modeled as the concatenation of cylindrical tubes by following the study of Ishizaka et al. (1976) and using morphological data (Weibel, 1963). The cross-sectional area is relatively constant along the section of about 20 cm below the glottis but increases rapidly in the remainder of the conductive zone and in the transitory and respiratory zones. The input impedance, Z_0 , is expressed as

$$Z_0 = -\frac{A_s Z_p + B_s}{C_s Z_p + D_s}, \quad (33)$$

with the matrix chaining given by

$$\begin{pmatrix} A_s & B_s \\ C_s & D_s \end{pmatrix} = \begin{pmatrix} A_{N_s} & B_{N_s} \\ C_{N_s} & D_{N_s} \end{pmatrix} \begin{pmatrix} A_{N_s-1} & B_{N_s-1} \\ C_{N_s-1} & D_{N_s-1} \end{pmatrix} \cdots \begin{pmatrix} A_1 & B_1 \\ C_1 & D_1 \end{pmatrix}, \quad (34)$$

where the first section corresponds to the lungs and the N_s th section to the portion just below the glottis. Z_p is the terminal impedance and is modeled as a capacity corresponding to an air volume of 500 ml (van den Berg, 1960).

The relationship between the Fourier transform of the acoustic pressure at the lip opening, P_r , and U_g can be written as

$$H = \frac{P_r}{U_g} = \frac{Z_r}{C_v Z_r - A_v} \cdot \frac{B_g - A_g Z_0}{Z_D} \quad (35)$$

based on Eq. (27) of Sondhi and Schroeter (1987), where Z_1 and Z_2 in the literature are replaced by Z_g and Z_1 , respectively.

D. Analysis of flow with low Reynolds numbers

The boundary-layer approximation is in general effective for flow with high Reynolds numbers. When the slit of the glottal opening is very narrow, the expected volume flow and Reynolds number may be small (Deverge *et al.*, 2003; Kaburagi and Tanabe, 2009). Flow in the glottal channel can then be treated as entirely viscous, and a model of fully developed Poiseuille flow (Pelorson *et al.*, 1994) can be adopted instead of the boundary-layer approximation. Using this model,

$$\Delta p = \frac{12\mu}{l_g} u_g \int_{x_0}^{x_e} \frac{1}{h(x)^3} dx, \quad (36)$$

where Δp is the pressure difference across the glottis and μ is the dynamic viscous coefficient.

Similar to Eqs. (6)–(8), a discrete-time representation can be obtained as

$$p_{F0} + \sum_{k=0}^{K-1} z_{\Delta}(k) u_g(n-k) = \frac{12\mu}{l_g} u_g(n) \int_{x_0}^{x_e} \frac{1}{h(x)^3} dx \quad (37)$$

and the volume flow can be estimated as

$$u_g(n) = \frac{p_{F0} + \sum_{k=1}^{K-1} z_{\Delta}(k) u_g(n-k)}{\frac{12\mu}{l_g} \int_{x_0}^{x_e} \frac{1}{h(x)^3} dx - z_{\Delta}(0)}. \quad (38)$$

In addition, the mean pressure can be estimated as

$$p(x) = p_{F0} - \frac{12\mu}{l_g} u_g \int_{x_0}^x \frac{1}{h(\chi)^3} d\chi + z_M(t) * u_g(t)|_{t=t_c} \quad (39)$$

instead of Eq. (16). The driving force for the lower mass can then be expressed as

$$f_1 = \lambda l_g \int_{x_0}^{x_e} p(x) dx. \quad (40)$$

E. Simulation procedure

The simulation procedure for the open-glottis condition is depicted in Fig. 3. The masses of the vocal-fold model are initially positioned in their rest states. The acoustic transmission matrices and input impedances are calculated for the sub- and supra-glottal tracts and the glottis using Eq. (20) and Eqs. (31)–(34) for the given area functions. The difference and mean acoustic pressures near the glottis are then determined [Eqs. (29) and (30)]. Next, the volume flow estimation and interactive boundary-layer method are alternately applied to determine the glottal flow and pressure distribution. The force driving the vocal-fold model [Eq. (17)] is calculated afterward, and the mechanical equations [Eqs. (13) and (14)] are solved numerically using the Runge–Kutta method. This procedure is repeated over the desired time steps.

The flow analysis [Eqs. (7)–(11)] is performed as follows. As the glottis opens, the channel takes a convergent shape, resulting in a flow separation point that is relatively constant near the glottal outlet. Therefore, the initial value of the separation point and the boundary-layer thickness can be set to their estimated values from the previous time step. The effective channel area in Eq. (2) is then determined from these initial values, and the volume flow is estimated from Eqs. (7) and (8). The interactive boundary-layer analysis follows and refines the estimation of the flow separation point and the boundary-layer thickness. The process is repeated two or three times to obtain the final estimation as the estimated value of the flow rate converges. As the glottis closes, the channel takes a divergent shape and the flow separation point varies significantly. Therefore, the separation point is initially set at the point where the glottal height is the minimum, while the boundary-layer thickness is set to zero.

In addition, the method used for flow analysis depends on the flow Reynolds number, $R = u_g l_g / \nu$. When the glottis opens, the analysis is first performed using the Poiseuille flow model of Sec. II D. As the simulation proceeds, if R of the previous time step is higher than a threshold, R_{thr} , the flow analysis method is switched to the interactive boundary-layer method. When the glottis closes, on the other hand, the flow Reynolds number typically decreases in time. The analysis is therefore first performed using the boundary-layer method but switched to the Poiseuille flow model if the resulting Reynolds number becomes smaller than R_{thr} . When the glottis is closed, the volume flow is set to zero.

III. NUMERICAL RESULTS

A. Frequency dependency of the source-filter coupling

Source-filter coupling is considered in our model as acoustic feedback from the sub- and supra-glottal tracts to the voice production system. The feedback is represented in the frequency domain in Eqs. (29) and (30). Therefore, these frequency characteristics are investigated using the vocal-tract area functions obtained by Story *et al.* (1996) and the sub-glottal area function obtained by Weibel (1963). Figure 4 shows the calculated results for the vowel /i/. The top plot is

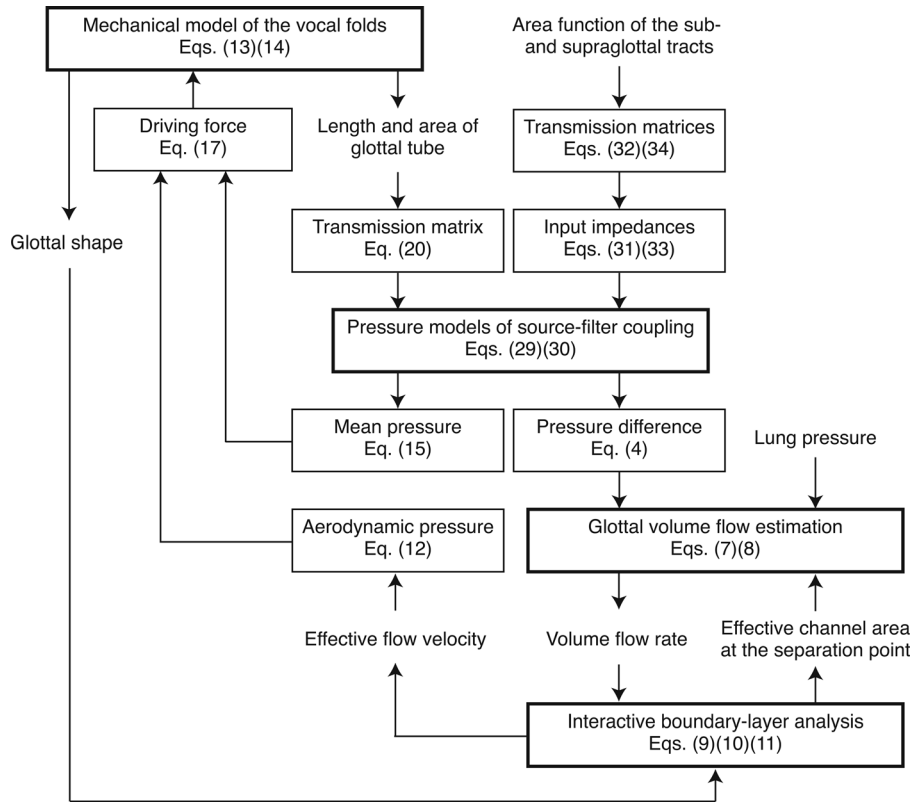


FIG. 3. The flow graph showing the simulation procedure for time-domain voice production.

the magnitude of the input impedance of the vocal tract, Z_1 . The magnitude response of the acoustic pressure difference, $\Delta P_A/U_g$, and the mean pressure, \bar{P}_A/U_g , are shown in Fig. 4(B) by solid and broken lines, respectively. Here, the length of the glottal tube (depth of the glottis) was set to $L_g = 0.3$ cm and the cross-sectional area to $S_g = 0.014$ cm². The air density was set to $\rho = 0.001184$ g/cm³ and the speed of sound to $c = 34630$ cm/s. Note that the results were obtained up to 10 kHz, even though plane wave propagation was assumed.

The results show that each frequency response has significant peaks around 4–5 kHz, corresponding to the fourth and fifth formants. Interestingly, these peaks are much greater than those corresponding to lower-order formants. It is also clear that $|\Delta P_A/U_g|$ and $|\bar{P}_A/U_g|$ are very similar in shape, with matching peak frequencies. This is because the acoustic pressure at the glottal inlet is generally much lower than that at the outlet. It indicates that $|P_{A0}| \ll |P_{A1}|$, and hence, $|\bar{P}_A/U_g|$ becomes approximately half of $|\Delta P_A/U_g|$, from their definitions. In Fig. 4(C), the magnitude responses are averaged for nine vowels based on area data of the vocal tract (Story *et al.*, 1996). We observe that several peaks are located around 3–5 kHz, with the largest peak just below 4 kHz, suggesting the importance of such high-frequency components in the source-filter coupling.

To examine the possible cause of these frequency-domain features, the input impedance of the vocal tract was calculated using a two-tube model. The results are shown in Fig. 5. The length and cross-sectional area of the epiglottal tube, L_1 and S_1 , respectively, were varied in this experiment, while the length of the secondary tube was set to $L_2 = 17 - L_1$ cm and

the area to $S_2 = 4$ cm². The peak frequency of the magnitude response is approximately 5.7 kHz for $L_1 = 1.5$ cm, 4.5 kHz for $L_1 = 2.0$ cm, and 3.5 kHz for $L_1 = 2.5$ cm. Each of these frequencies corresponds to the fundamental resonant frequency, $f = c/(4L_1)$, of a uniform acoustic tube with one-end open and one-end closed. The figure also shows that the magnitude varies significantly with cross-sectional area, S_1 . Figures 4 and 5 support the result of Titze (2008) in showing that the source-filter coupling can be enhanced by reducing the epiglottal cross-sectional area. This occurs because a peak in the input impedance forms a peak in the magnitude responses, $|\Delta P_A/U_g|$ and $|\bar{P}_A/U_g|$. The length, $L_1 = 1.5$ or 2 cm seems to be typical for the epiglottal section, and the mean cross-sectional area within this section is approximately 0.25 cm² for vowels (Story *et al.*, 1996).

Figure 6 shows the amplitude and phase responses of the pressure difference and mean pressure for the vowel /i/ within the frequency range of 0–800 Hz. The cross-sectional area of the glottal tube was varied over $S_g = 0.014$, 0.028, and 0.14 cm², while the length was fixed at $L_g = 0.3$ cm. Equations (29) and (30) each form a frequency response of a linear filter. Hence, each frequency component of U_g , the Fourier transform of the volume flow input to these filters, produces the pressure difference or mean pressure of the same frequency. The magnitude response can be interpreted as the frequency-dependent strength of the source-filter coupling because it determines the amplitude of the resulting pressure values. On the other hand, the phase response is related to the type of coupling because it determines the phase difference of each frequency component between the input volume flow and the output pressure value.

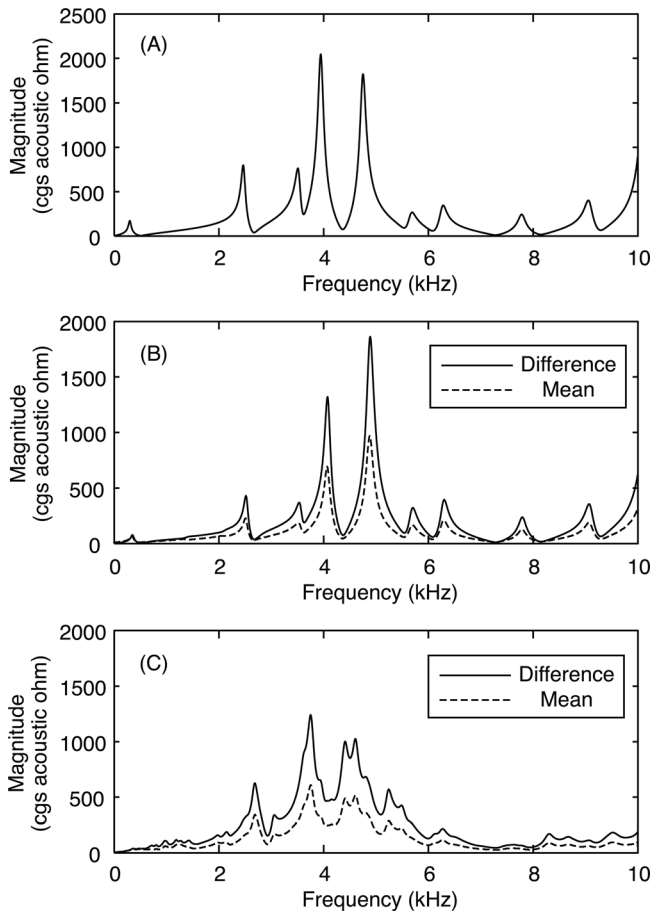


FIG. 4. Computed results of (A) vocal-tract input impedance and (B) frequency responses of the acoustic pressure difference (solid) and mean acoustic pressure (broken) across the glottis for the vowel /i/. Mean frequency response over nine vowels is plotted in (C) for the acoustic pressure difference and mean acoustic pressure, respectively.

Regarding the pressure difference shown in the left column of Fig. 6, suppose that the frequency, f_c , of a component of U_g is smaller than the peak frequency, f_p , of the frequency response (this peak frequency corresponds to the first formant of the vowel and is approximately 350 Hz here). The phase is approximately $-\pi/2$ in this frequency region, but it approaches $-\pi$ as f_c approaches f_p . The temporal relationship between a frequency component of the glottal flow and the equivalent component of the pressure difference is illustrated in Fig. 7, where both components are depicted as unit-amplitude sinusoids with the same frequency by ignoring the magnitude response. Suppose that the phase response is $-\pi/2$ at that frequency. The pressure difference, $\Delta p_A = p_{A0} - p_{A1}$, is negative when the flow increases from zero and produces a reverse flow at that moment. On the other hand, the pressure difference is positive when the flow decreases, causing a forward flow that increases the flow rate. These observations imply that the pressure difference across the glottis can skew the glottal flow. Supposing instead the phase response is $-\pi$, it is clear from Fig. 7(B) that the flow generated by the acoustic pressure difference tends to cancel the glottal flow. Note, however, that the degree of this flow cancelation depends on the magnitude response at that frequency. The coupling effect mentioned above can independently be

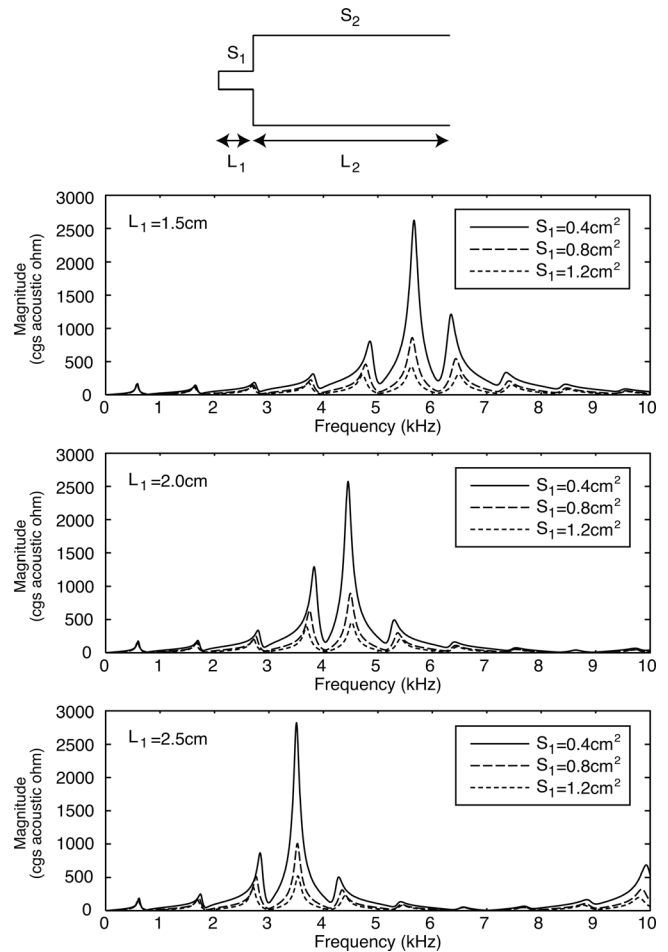


FIG. 5. Input impedance of a two-tube model resembling the narrow epiglottal section and the main section of the vocal tract. Results are compared for a number of epiglottal section lengths, L_1 , and cross-sectional areas, S_1 .

applied to each frequency component of the glottal flow. Therefore, if the frequency of a component is below f_p and that of another one slightly above f_p , the feedback pressure difference can skew the first component and degrade the second one.

Returning to Fig. 6 and considering the frequency responses related to the mean acoustic pressure, the phase response is positive when the frequency is below f_p and negative when the frequency exceeds f_p . In Fig. 7, the phase difference between the two sinusoids is set to $\pi/4$ in Fig. 7(C) and $-\pi/4$ in Fig. 7(D) as the typical value for each of these cases. Here, the volume flow is the input and the pressure difference is the output of the filter, \bar{p}_A/U_g . Because the volume flow is approximately proportional to the glottal area, the waveform pattern of u_g is similar to that of the vocal-fold displacement, especially when the frequency component in interest is the fundamental. In Fig. 7(C), the positive peak of the mean pressure occurs when the glottis begins to open. The pressure then decreases and reaches a negative peak, whereupon the glottis may be closing. Thus, the acoustic feedback is favorable to the maintenance of the vocal-fold oscillation. In Fig. 7(D), on the contrary, the positive peak of the pressure occurs during adductive motion of the vocal-fold, thus preventing oscillation. The plots in Fig. 6 also

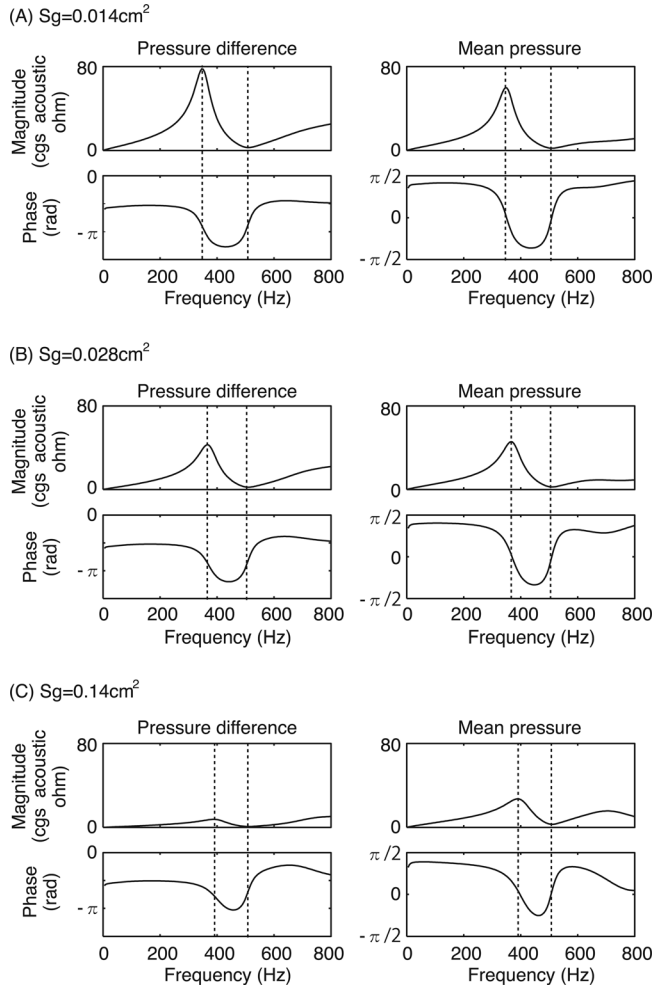


FIG. 6. Frequency responses of the acoustic pressure difference and mean acoustic pressure for the vowel /i/. The cross-sectional area of the uniform tube for the glottal section in Fig. 2 was changed to (A) 0.014 cm, (B) 0.028 cm, and (C) 0.14 cm. Both magnitude and phase responses are shown. The broken vertical lines indicate the frequencies of the first peak and first dip.

shows that the peak frequency, f_p , increases as the cross-sectional area of the glottal tube increases. In addition, the magnitude of every frequency response generally decreases as the glottis opens, suggesting that the strength of the source-filter coupling can vary during the glottal cycle.

B. Time-domain simulation

Next, following the procedure in Sec. II E, our voice production model was used to simulate vibration of the vocal folds and generation of the glottal sound source. Parameters of the vocal-fold model (Pelorson *et al.*, 1994) were set to $m_1 = 0.125/Q$ g, $m_2 = 0.025/Q$ g, $k_1 = 80000Q$ dyn/cm, $k_2 = 8000Q$ dyn/cm, $k_{12} = 55000Q$ dyn/cm, $r_1 = 0.2\sqrt{m_1 k_1} = 23.3$ dyn s/cm, and $r_2 = 1.2\sqrt{m_2 k_2} = 18.6$ dyn s/cm, where Q is the tension parameter. For the glottis in a closed state, the values of the stiffness parameters were increased such that $k_1 = 320000Q$ and $k_2 = 32000Q$ dyn/cm. The parameters of the dampers were also increased, such that $r_1 = 257$ and $r_2 = 49.6$ dyn s/cm. The initial displacement was set to $y_0 = 0.00014$ cm and the vocal-fold length to $l_g = 1.2$ cm. λ in Eq. (17) was determined so that the effective depth of the lower mass,

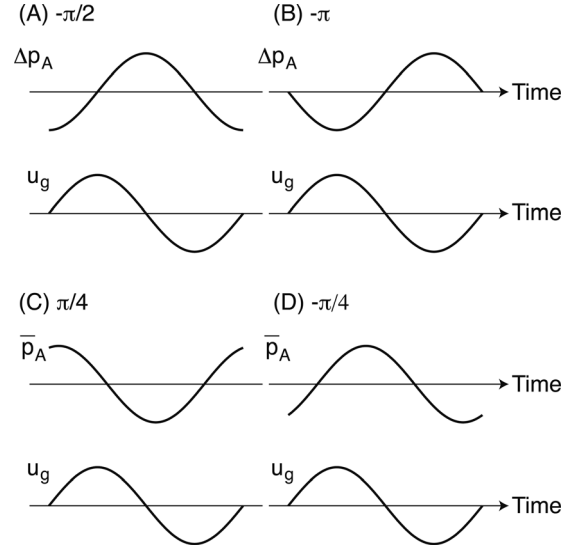


FIG. 7. Waveform plots of sinusoids illustrating the input output relation of the acoustic pressure difference and mean acoustic pressure frequency responses. The volume flow and specific pressure value are the input and output, respectively, of the linear, causal filters defined by these responses. The effect of the source-filter coupling can be estimated by the phase difference, which specifies the interrelation of these variables. Here, the influence of the magnitude response is not considered.

$\lambda(x_e - x_0)$, was about 0.27 cm. The threshold Reynolds number, R_{thr} , was 200. The dynamic viscous coefficient was set to $\mu = 0.0182 \times 10^{-3}$ Pa s, and the sub-glottal pressure was set to $p_{F0} = 8$ cm H₂O. A sampling frequency of 20 kHz was used to perform the simulation, and therefore the time step was 0.05 ms.

Figure 8 shows the temporal pattern of the main physical variables, calculated for the vowel /i/, with a tension parameter value of $Q = 1$. From top to bottom, the plots show glottal volume flow (u_g), acoustic pressure difference across the glottis (Δp_A), mean acoustic pressure (\bar{p}_A), mass displacements (y_1 and y_2), driving force of the lower mass (f_1), the point of flow separation relative to the thickness of the vocal fold (x_s/x_c), and the displacement thickness of the boundary layer relative to the channel height at the flow separation point ($2\delta(x_s)/h(x_s)$). The displacements for the lower and upper masses are shown by the solid and broken lines, respectively. The aerodynamic component of the driving force is shown by the broken line, while the solid line represents the sum of both aerodynamic and acoustic forces.

The pressure difference is generally positive and a peak is observed before the glottal closure, resulting in an increase in the glottal flow. The mean pressure is negative before the closure and has the effect of pulling both folds medially. The results also suggest that the vocal-fold vibration is maintained by the positive-to-negative change of the driving force, in which the aerodynamic component is principal. The point of flow separation was near the glottal exit (x_e) during the opening phase and it gradually moved in the upstream direction during the closing phase, in accordance with the convergent-divergent change in the glottal shape. Lastly, the bottom plot of the displacement thickness indicates the rate of channel height reduction, because the effective channel height is given as $h(x_s) - 2\delta(x_s)$. This channel height

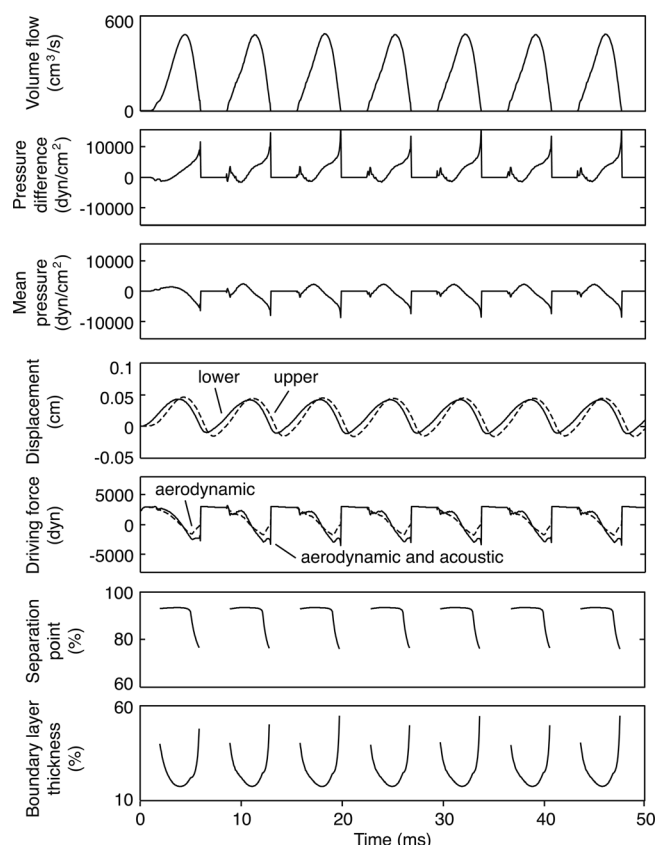


FIG. 8. The results of a time-domain simulation of voice production for the vowel /i/. The plots show the changes with time of the glottal volume flow, acoustic pressure difference across the glottis, mean acoustic pressure in the glottis, mass displacements, driving force on the vocal folds, flow separation point, and boundary-layer thickness at the separation point. The displacements of the lower and upper masses are shown by solid and broken lines, respectively. The broken line in the driving force plot shows the force component generated by the aerodynamic pressure. The flow separation point is normalized by the thickness of the vocal fold. The boundary-layer thickness is normalized by the nominal height of the glottal channel at the separation point.

reduction almost exceeded 20% in this simulation. The influence of the boundary layer was especially significant when the glottis opened or closed, at which time the volume flow rate and hence the Reynolds number were small.

Figure 9 depicts the shape of the medial portion of the vocal fold (solid line) and the outline of the boundary layer formed by the displacement thickness (broken line) for six time instants of the simulation shown in Fig. 8. The arrow indicates the point of flow separation on the surface of the vocal fold. As time proceeds from 2 to 4 ms, the glottal shape changes from being convergent to being uniform. The channel height widens and the relative value of the boundary-layer thickness, $2\delta(x_s)/h(x_s)$, decreases. At 5 and 5.5 ms, the channel has a divergent shape, the separation point shifts upstream, and the boundary-layer thickness increases again. These features of the glottal shape change and the behavior of the boundary layer are typical (Pelorson *et al.*, 1994; Story and Titze, 1995; Kaburagi and Tanabe, 2009), but it is worth pointing out again that the development of the boundary layer can significantly reduce the glottal channel and affect the estimation of the glottal flow based on the pressure-flow

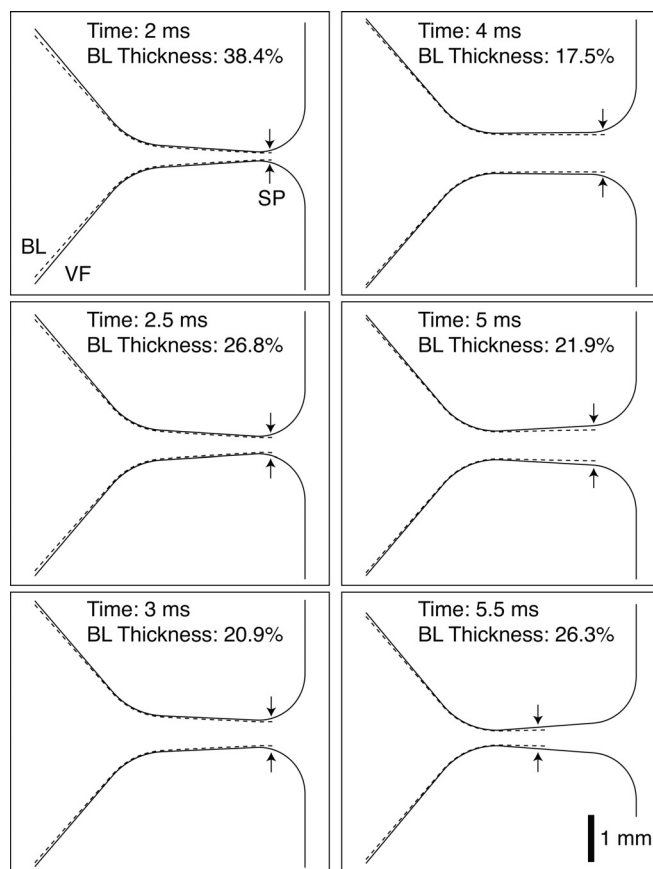


FIG. 9. The shape of the vocal folds and the displacement thickness of the boundary layer obtained at six time points of the simulation plotted in Fig. 8. Left is toward the sub-glottal region and right toward the vocal tract. VF denotes the vocal fold, SP the flow separation point on the surface of the vocal fold, and BL the boundary layer. The numerical value of the boundary-layer thickness is normalized by the nominal height of the glottal channel at the flow separation point.

relation [see Eq. (1)]. Also note that the boundary-layer analysis is basically effective upstream of the separation point.

Next, the contribution of the source-filter coupling was examined by changing the frequency responses of the acoustic pressures. The responses, $\Delta P_A/U_g$ and \bar{P}_A/U_g , for the vowel /i/ were low-pass filtered with a cut-off frequency of 3.8 kHz at each time step of the simulation using a zero-phase digital filter. The filtered versions of the responses were then used with Eqs. (4) and (15) to calculate the pressure difference and mean pressure. An example filtered magnitude response is shown in the upper part of Fig. 10 for a glottal area of 0.014 cm². By comparing Figs. 4(B) and 10, we can see that the significant peaks around 4 and 5 kHz are removed by filtering. The traces in Fig. 10 show the results of a voice production simulation. The filtering of the frequency responses degrades the strength of the source-filter coupling. With the absence of high-frequency peaks, the vocal-fold oscillation is attenuated quickly within a few cycles, suggesting that these peaks are indispensable.

Finally, the tension parameter of the vocal fold, Q , was changed from two to five linearly with time. Figure 11 shows the spectrogram of the glottal flow, u_g , within the frequency range 0–2 kHz. The horizontal broken lines around 350 and 500 Hz correspond to the first peak and dip, respectively, of

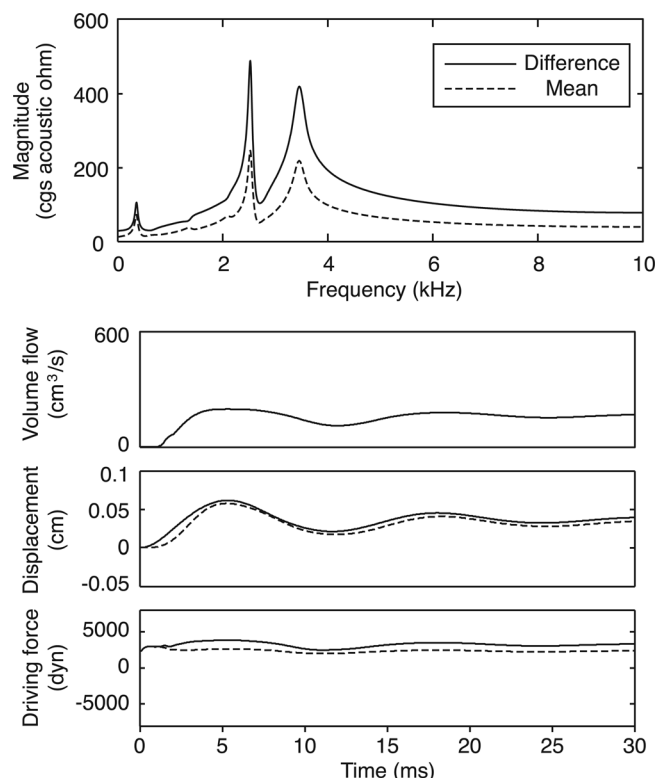


FIG. 10. The frequency plot shows the filtered responses of the acoustic pressure difference (solid) and mean acoustic pressure (broken) for the vowel /i/. Time traces show simulation results obtained using these responses. The details of the broken lines are the same as that given in Fig. 8.

$\Delta P_A/U_g$ and \bar{P}_A/U_g . The frequency region enclosed by the broken lines is therefore unfavorable for phonation (Ishizaka and Flanagan, 1972; Titze, 2008). The fundamental frequency reaches this peak frequency at a time of approximately 0.15 s. Between 0.15 and 0.28 s, the spectrogram indicates the appearance of subharmonic and unstable behavior. The harmonic pattern then stabilizes when, at 0.28 s, the fundamental frequency agrees with the dip frequency of approximately 500 Hz.

Figure 11 also shows waveforms of the main variables for the same simulation. In the first time section (0.02–0.04 s), the fundamental frequency is outside the unfavorable frequency region. Skew in the glottal flow waveform and a stable positive-to-negative change of the driving force are observed. In the second section (0.10–0.12 s), the fundamental frequency is closer to the unfavorable region. The pressure difference waveform exhibits sharp positive peak each time the glottis opens or closes, and the mean pressure has a negative peak at each of these moments. In the 0.20–0.22 s interval, where the spectrogram shows the existence of subharmonic behavior, there are three broad peaks in the pressure difference waveform, as indicated by arrows. By comparison with the waveform of 0.10–0.12 s, it might be suggested that these broad peaks are generated by a merging of the peaks, appearing in 0.10–0.12 s, between the adjacent glottal cycles. Finally, over 0.30–0.32 s, regular vocal-fold oscillation is recovered. Note that the fundamental frequency is close to the dip frequency of $\Delta P_A/U_g$ and \bar{P}_A/U_g . The feedback acoustic pressure is therefore weak.

Figure 12 shows spectrogram and waveform plots obtained by a simulation in which the frequency responses were low-pass filtered, as for the experiment plotted in Fig. 10. When the tension parameter is changed from two to five, self-oscillation of the vocal folds is recovered irrespective of the filtering. Some differences are readily noticeable between the spectrograms in Figs. 11 and 12. When the high-frequency peaks are removed, the fundamental frequency is lower at the beginning of the simulation. After the unstable region about 0.16–0.28 s, the fundamental frequency remains below the first peak frequency, even though the tension increases with time. The recovery of the self-oscillation and the locking of the oscillation frequency are due to the enhancement of source-filter coupling that produces periodic change in the pressure near the glottis, as discussed in more detail in the subsequent section.

Significant differences are also observed between the waveform plots of Figs. 11 and 12. In the first time section, the amplitudes of the pressure difference and the mean pressure decrease significantly as shown in Fig. 12, resulting in the reduction of the amplitude of the flow and driving force waveforms. As time proceeds, the fundamental frequency approaches the first peak of the responses. The source-filter coupling is then enhanced and the waveforms of Figs. 11 and 12 are similar in the second and third time intervals, although the waveforms of the flow and pressure values are generally smoother in Fig. 12. This is due to the elimination of the high-frequency components of the acoustic feedback by the filtering operation. Finally, in the fourth section, the pressure waveforms are much more prominent in Fig. 12 because the fundamental frequency remains near the peak region of the responses. The vocal folds did not oscillate when the value of the tension parameter is increased above one because they are separated after several cycles (see Fig. 10).

IV. SUMMARY AND DISCUSSION

In this work, a model was presented to explain the self-oscillation of vocal folds and the periodic changes of glottal flow. The model was constructed with an accurate flow analysis method based on a boundary-layer approximation that gives the effective flow channel by considering the thickness of the boundary layer. The variable flow separation point was also used to estimate the volume flow precisely. The generated acoustic waves travel through the vocal and subglottal tracts, resulting in the formation of an acoustic field near the glottis. This acoustic feedback was represented in terms of the pressure difference across the glottis, which contributes to the volume flow generation, and the mean pressure in the glottis, which partly drives the vocal folds. These specific pressure values were expressed in the frequency domain using linear, causal filters. Their frequency responses gave explicit relations between glottal flow and the resulting pressure value for each frequency.

Besides the explicit use of the interactive boundary-layer analysis, our voice production model is slightly different from that studied by Titze (2008) in modeling the acoustic wave propagation through the glottis. Equations (1) and

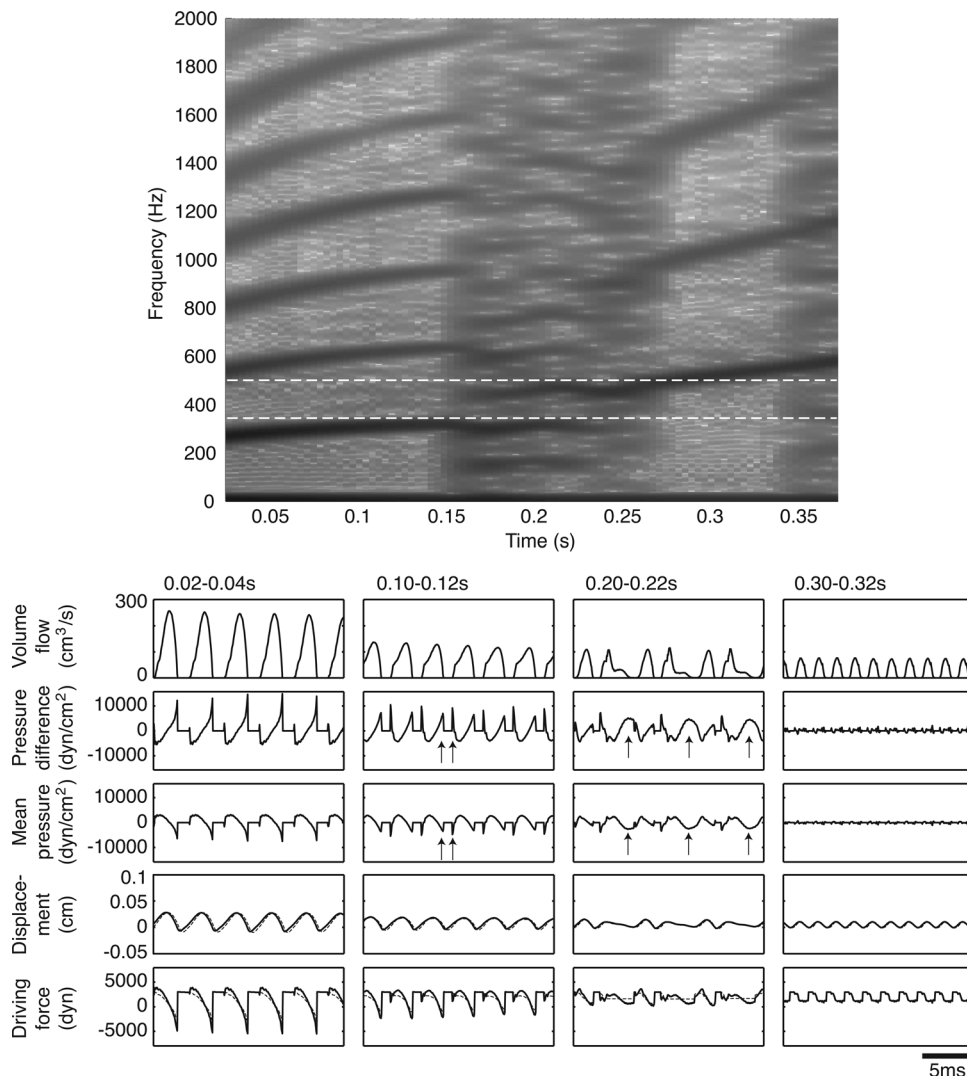


FIG. 11. The spectrogram for the temporal pattern of the glottal volume flow. The tension parameter of the vocal-fold model was linearly changed in time from two to five. A Hamming window with a length of 1024 samples was used. The shift width of the window was 950 samples. Traces below the spectrogram show the simulated waveforms of the main physical variables. Each time section corresponds to a portion of the spectrogram. The details of the broken lines in the displacement and driving force waveforms are the same as that given in Fig. 8. Arrows in the second and third time sections indicate characteristic peak patterns of the pressure values.

(2) of the study by Titze (2008) indicate that the wave propagation through the glottis is independent of the cross-sectional area. The glottal area is incorporated as the dominant parameter of the volume flow estimation using the pressure-flow relation given in Eq. (4) of Titze's study. Note that the input impedance, denoted by Z_1 and Z_2 , can be determined from the area function of the sub- or supra-glottal tracts and an appropriate terminal impedance. At this point, the flow equation [Eq. (5) of Titze's study] only includes the entrance areas, A_s and A_e , as the information of these tracts, but the forward and backward waves, p_s^+ and p_e^- , reflect implicitly their acoustic properties. In our model, the flow estimation [Eq. (1)] is based on the effective glottal area measured at the estimated flow separation point by excluding the estimated thickness of the boundary layer. In addition, the glottal area and its wave-traveling length are considered as the parameter determining the acoustic pressure above and below the glottis [Eqs. (27) and (28)]. As a result, the specific pressures representing the source-filter coupling (i.e., the pressure difference and mean pressure of the glottis) are dependent on the glottal area, implying that the strength of the coupling effect varies during the glottal cycle.

Numerical investigations were performed to examine the effect of the source-filter coupling. First, the input impedance of the vocal tract and the frequency responses of the pressure difference and mean pressure were calculated. They showed significant peaks in the frequency range 3–5 kHz. Note that the results were obtained by using the acoustic tube model, which assumes plane wave propagation, and therefore they would be less accurate in frequency ranges higher than approximately 4 kHz. Although our method has this limitation, we suggest the acoustic resonance of the epiglottal tube is the cause of the high-frequency peaks, in agreement with the statement of Titze (2008) that the source-filter coupling can be enhanced by reducing the epiglottal cross-sectional area. A more direct relationship between the laryngeal cavity and high-order formants was investigated by Takemoto *et al.* (2006), with results that support our own. Using magnetic resonance imaging and a transmission line model of the vocal tract, they showed that the acoustic resonance of the vestibular and ventricular spaces determines the fourth formant independent of the section of vocal tract above the laryngeal cavity. Concerning the acoustics of the vocal tract, further studies would address

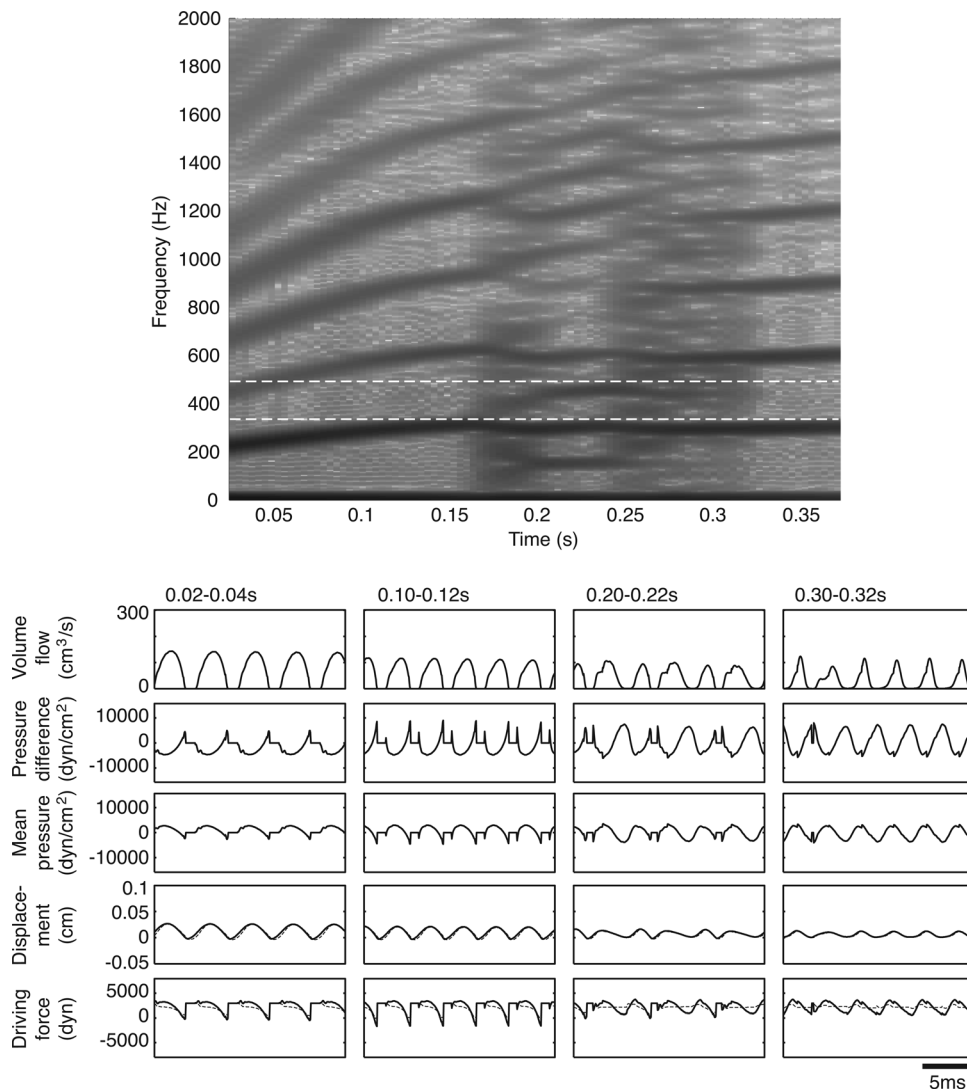


FIG. 12. The spectrogram for the temporal pattern of the glottal volume flow when the frequency responses of the specific pressure values were low-pass filtered to eliminate the high-frequency peaks, as shown in the upper part of Fig. 10. The tension parameter of the vocal-fold model was changed in the same way as in Fig. 11. Traces show the simulated waveforms of the main physical variables. Each time section corresponds to a portion of the spectrogram. The details of the broken lines in the displacement and driving force waveforms are the same as that given in Fig. 8.

the influence of a three-dimensional acoustic field in the tract and of the piriform fossae, which causes antiresonances in the vocal-tract transfer function in the frequency range 4–5 kHz (Dang and Honda, 1997).

The importance of these high-frequency peaks was confirmed using time-domain simulations. When the significant peaks were eliminated by a zero-phase low-pass filter and the resulting frequency responses were used to calculate the specific pressure values near the glottis, the vocal-fold oscillation was attenuated quickly, resulting in a convergence of the mass displacement. The resulting driving force was almost constant and movements of both masses were synchronized, indicating that convergent–divergent changes in the glottal channel were not attained. Elimination of the dominant high-frequency peaks led to the energy reduction of the acoustic feedback. Foregoing studies, Zañartu *et al.* (2007), for example, showed that the vocal-fold oscillation was not attained when the sub- and supra-glottal tracts had infinite length and there was no acoustic reflection from the tracts. This can be regarded as an extreme case of our filtering experiment, although some simulation studies reported that the vocal-fold model can oscillate without the source-fil-

ter coupling (Tokuda *et al.*, 2007; Titze, 2008). We also showed that the vocal-fold oscillation was recovered by increasing the tension parameter of the vocal-fold model because the fundamental frequency approached the first formant and the source-filter coupling was enhanced.

The results of time-domain simulations with a variable tension parameter showed bifurcation and voice instability (Tokuda *et al.*, 2007). Detailed analysis was performed on the specific patterns appearing in the pressure difference and mean pressure waveforms, by paying attention to the interrelation between the fundamental frequency and the first peak and dip of the pressure frequency responses.

When the high-frequency peaks appearing in the frequency responses of the pressure values were removed in the variable fold-tension simulation, the result was quite different. The most interesting point was that the fundamental frequency remained below the first peak frequency of the responses, corresponding to the first formant, even though the tension of the vocal fold changed significantly. This peculiar phenomenon can be attributed to a strong periodicity generated in the pressure difference and mean pressure waveforms. Because the amplitude and phase difference of

the vocal-fold oscillation generally decreased as the tension parameter increased, the temporal change of the aerodynamic force also weakened, and instead, the influence of the acoustically oriented driving force dominated the self-oscillation (see the fourth time section of Fig. 12). The fundamental frequency was therefore controlled by the changing the rate of the mean acoustic pressure, which was very close to the first peak frequency of approximately 350 Hz.

From the simulation experiments, we conclude that the elimination of the high-frequency components of the source-filter coupling can lead to an enhancement of the influence of low-frequency peaks and change the consequences of the coupling. Future studies will aim at clarifying further the meaning of the frequency-dependent characteristics of the source-filter coupling in the process of human voice production and singing.

ACKNOWLEDGMENT

This research was partly supported by the Grant-in-Aid for Scientific Research from the Japan Society for the Promotion of Science (JSPS) (Grant Nos. 18500134 and 19103003).

- Berry, D. A., and Titze, I. R. (1996). "Normal modes in a continuum model of vocal fold tissues," *J. Acoust. Soc. Am.* **100**, 3345–3354.
- Broad, D. J. (1979). "The new theories of vocal fold vibration," in *Speech and Language: Advances in Basic Research and Practice*, vol. 2, edited by N. J. Lass (Academic Press, New York), pp. 203–256.
- Chan, R. W., and Titze, I. R. (2006). "Dependence of phonation threshold pressure on vocal tract acoustics and vocal fold tissue mechanics," *J. Acoust. Soc. Am.* **119**, 2351–2362.
- Dang, J., and Honda, K. (1997). "Acoustic characteristics of the piriform fossa in models and humans," *J. Acoust. Soc. Am.* **101**, 456–465.
- Deverge, M., Pelorson, X., Vilain, C., Lagrée, P.-Y., Chentouf, F., Willems, J., and Hirschberg, A. (2003). "Influence of collision on the flow through in-vitro rigid models of the vocal folds," *J. Acoust. Soc. Am.* **114**, 3354–3362.
- Flanagan, J. L. (1972). *Speech Analysis Synthesis and Perception*, 2nd ed. (Springer Verlag, New York), pp. 36–38.
- Hofmans, G. C., Groot, G., Ranucci, M., Graziani, G., and Hirschberg, A. (2003). "Unsteady flow through in-vitro models of the glottis," *J. Acoust. Soc. Am.* **113**, 1658–1675.
- Ishizaka, K., and Flanagan, J. L. (1972). "Synthesis of voiced sounds from a two-mass model of the vocal cords," *Bell Syst. Tech. J.* **51**, 1233–1268.
- Ishizaka, K., Matsudaira, M., and Kaneko, T. (1976). "Input acoustic-impedance measurement of the subglottal system," *J. Acoust. Soc. Am.* **60**, 190–197.
- Kaburagi, T. (2008). "On the viscous-inviscid interaction of the flow passing through the glottis," *Acoust. Sci. Technol.* **29**, 167–175.
- Kaburagi, T., and Tanabe, Y. (2009). "Low-dimensional models of the glottal flow incorporating viscous-inviscid interaction," *J. Acoust. Soc. Am.* **125**, 391–404.
- Kalse, S. G. C., Bijl, H., and van Oudheusden, B. W. (2003). "A one-dimensional viscous-inviscid strong interaction model for flow in indented channels with separation and reattachment," *ASME J. Biomech. Eng.* **125**, 355–362.
- Lous, N. J. C., Hofmans, G. C., Veldhuis, R. N. J., and Hirschberg, A. (1998). "A symmetrical two-mass vocal-fold model coupled to vocal tract and trachea, with application to prosthesis design," *Acust. Acta Acust.* **84**, 1135–1150.
- Pelorson, X., Hirschberg, A., van Hassel, R. R., Wijnands, A. P. J., and Aurégan, Y. (1994). "Theoretical and experimental study of quasi steady-flow separation within the glottis during phonation. Application to a modified two-mass model," *J. Acoust. Soc. Am.* **96**, 3416–3431.
- Rothenberg, M. (1981). "Acoustic interaction between the glottal source and the vocal tract," in *Vocal Fold Physiology*, edited by K. N. Stevens and M. Hirano (University of Tokyo Press, Tokyo, Japan), pp. 305–328.
- Scherer, R. C., Titze, I. R., and Curtis, J. F. (1983). "Pressure-flow relationships in two models of the larynx having rectangular glottal shapes," *J. Acoust. Soc. Am.* **73**, 668–676.
- Schlichting, H., and Gersten, K. (2000). *Boundary-Layer Theory*, 8th ed. (Springer Verlag, New York), pp. 196–201.
- Sondhi, M. M., and Schroeter, J. (1987). "A hybrid time-frequency domain articulatory speech synthesizer," *IEEE Trans. Acoust., Speech, Signal Process.* **ASSP-35**, 955–967.
- Story, B. H., and Titze, I. R. (1995). "Voice simulation with a body-cover model of the vocal folds," *J. Acoust. Soc. Am.* **97**, 1249–1260.
- Story, B. H., Titze, I. R., and Hoffman, E. A. (1996). "Vocal tract area functions from magnetic resonance imaging," *J. Acoust. Soc. Am.* **100**, 537–554.
- Švec, J. G., Schutte, H. K., and Miller, D. G. (1999). "On pitch jumps between chest and falsetto registers in voice: Data from living and excised human larynges," *J. Acoust. Soc. Am.* **106**, 1523–1531.
- Takemoto, H., Adachi, S., Kitamura, T., Mokhtari, P., and Honda, K. (2006). "Acoustic roles of the laryngeal cavity in vocal tract resonance," *J. Acoust. Soc. Am.* **120**, 2228–2238.
- Titze, I. R. (1973). "The human vocal cords: A mathematical model, part I," *Phonetica* **28**, 129–170.
- Titze, I. R. (1988). "The physics of small-amplitude oscillation of the vocal folds," *J. Acoust. Soc. Am.* **83**, 1536–1552.
- Titze, I. R. (2008). "Nonlinear source-filter coupling in phonation: Theory," *J. Acoust. Soc. Am.* **123**, 2733–2749.
- Titze, I. R., and Worley, A. S. (2009). "Modeling source-filter interaction in belting and high-pitched operatic male singing," *J. Acoust. Soc. Am.* **126**, 1530–1540.
- Tokuda, I. T., Horáček, J., Švec, J. G., and Herzel, H. (2007). "Comparison of biomechanical modeling of register transitions and voice instabilities with excised larynx experiments," *J. Acoust. Soc. Am.* **122**, 519–531.
- van den Berg, J. (1960). "An electrical analogue of the trachea, lungs and tissues," *Acta Physiol. Pharmacol. Neerl.* **9**, 361–385.
- van den Berg, J., Zantema, J. T., and Doornenbal, P. (1957). "On the air resistance and the Bernoulli effect of the human larynx," *J. Acoust. Soc. Am.* **29**, 626–631.
- Weibel, E. R. (1963). *Morphometry of the Human Lung* (Springer, Berlin, Germany), pp. 136–140.
- Zañartu, M., Mongeau, L., and Wodicka, G. R. (2007). "Influence of acoustic loading on an effective single mass model of the vocal folds," *J. Acoust. Soc. Am.* **121**, 1119–1129.
- Zhang, Z., Neubauer, J., and Berry, D. A. (2006). "Aerodynamically and acoustically driven modes of vibration in a physical model of the vocal folds," *J. Acoust. Soc. Am.* **120**, 2841–2849.
- Zhao, W., Zhang, C., Frankel, S. H., and Mongeau, L. (2002). "Computational aeroacoustics of phonation, part I: Computational methods and sound generation mechanisms," *J. Acoust. Soc. Am.* **112**, 2134–2146.