

Voice production model employing an interactive boundary-layer analysis of glottal flow

Tokihiko Kaburagi¹, Katsunori Daimo², and Shogo Nakamura³

¹ Department of Acoustic Design, Faculty of Design, Kyushu University,

² Graduate School of Design, Kyushu University,

³ School of Design, Kyushu University,

4-9-1 Shiobaru, Minami-ku, Fukuoka, 815-8540 Japan

kabu@design.kyushu-u.ac.jp

Abstract

A voice production model has been studied by considering essential aerodynamic and acoustic phenomena in human phonation. Acoustic voice sources are produced by the temporal change of volume flow passing through the glottis. A precise flow analysis is therefore performed based on the boundary-layer approximation and the viscous-inviscid interaction between the boundary layer and core flow. This flow analysis can supply information on the separation point of the glottal flow and the thickness of the boundary layer, which strongly depend on the glottal configuration, and yield an effective prediction of the flow behavior. When the flow analysis is combined with a mechanical model of the vocal fold, 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 interaction. Preliminary simulations were conducted by changing the relationship between the fundamental and formant frequencies and their results were reported.

Index Terms: voice production, boundary-layer analysis, source-filter interaction

1. Introduction

Human voice production is a highly complicated process involving mutual interactions of the flow and vocal folds through flow-driven mechanical oscillations and shaping of the glottal channel. Air volume flow passing through the glottis, the acoustic source for voiced sounds, is also affected by the acoustic properties of the vocal tract. In addition, Titze showed, in his notion of second level interaction [1] that the acoustic property of the sub- and supra-glottic cavity can enhance or suppress vocal fold oscillations.

Based on the boundary-layer approximation [2], the first author studied an accurate method for predicting one- and two-dimensional behavior of glottal flow [3, 4]. The method considers the viscous-inviscid interaction, i.e., the mutual interactions between the viscous region of the glottal flow, called the boundary layer, and the inviscid flow region, called the core flow. The core flow velocity is then effectively estimated, and the momentum-integral equation of the boundary layer is solved for this effective flow velocity. Since the volume flow is nearly proportional to the effective sectional area of the glottal channel at the separation point, information on the layer thickness and the flow separation point is clearly essential for glottal flow estimation and voice production modeling. It is noteworthy that

the method is applicable for arbitrary smoothly-shaped glottal configurations.

This paper presents a voice generation model that can explain essential aerodynamic and acoustic phenomena in human phonation. Together with the interactive boundary-layer analysis of the glottal flow, an acoustic tube model [5] is employed for the supra-glottal region, and a mechanical model similar to the two-mass model [6] for the vocal folds. In combination, flow-driven vibration of the vocal folds produces periodic changes in the volume flow, and the resulting acoustic fluctuation propagates in the tract. To consider the aerodynamic-acoustic interaction [1], the acoustic pressure in the vicinity of the glottis is also used to determine the volume flow. Results are shown for a number of numerical experiments by varying the relationship between the fundamental frequency of the vocal fold vibration and the formant frequencies of the vocal tract.

2. Interactive boundary-layer analysis

By inputting the subglottal pressure and information about the shape of the glottal channel, the glottal flow analysis predicts the volume flow and pressure distribution along the channel. The temporal waveform of the volume flow serves as the acoustic source of voiced sounds, while the pressure of the expiratory flow mechanically drives the vocal folds. Our method is based on a boundary-layer approximation that includes the interaction between the core flow and the boundary layer [3, 4].

By assuming that the glottal flow is one dimensional, incompressible and quasi-steady [2], the boundary-layer equation of von Kármán is solved by using the Pohlhausen method [7]. The characteristic quantities, such as the displacement thickness, momentum thickness, and wall shear stress, are then estimated together with the point of flow separation, which varies significantly according to the glottal shape. The boundary-layer equation takes the velocity of the core flow as the boundary condition. However, the core flow velocity is in turn affected by the development of the boundary layer that reduces the effective size of the channel. This indicates that the equations of the core flow and the boundary layer should be coupled together for an effective analysis of the glottal flow [3, 4].

2.1. Effective velocity of the core flow

As illustrated in Fig. 1, the glottal channel is assumed to be symmetrical with respect to the x -axis. The height parameter $h(x)$ represents the distance between both vocal folds, while L_g denotes the length of the fold. Assuming incom-

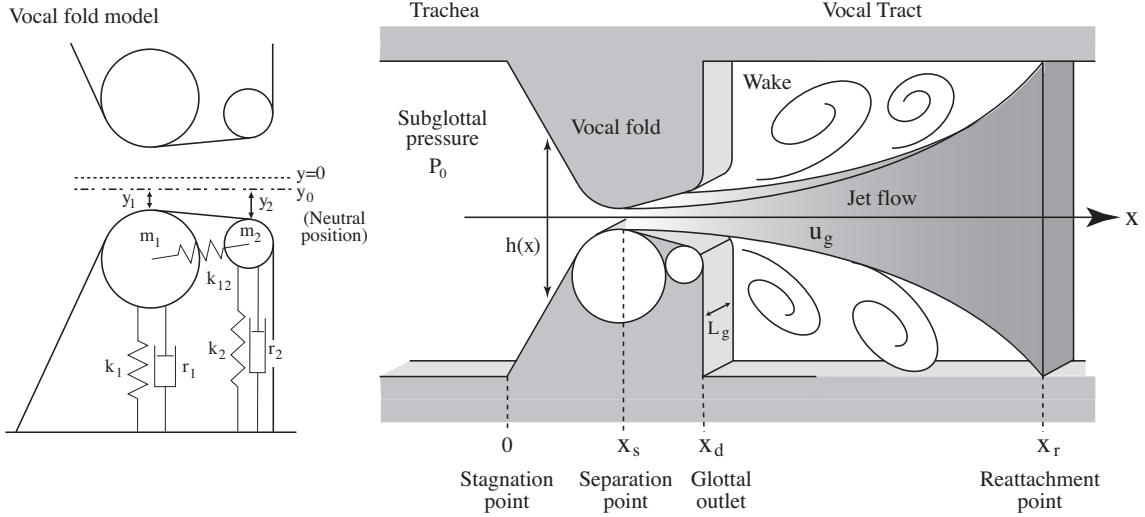


Figure 1: Model of the symmetrical glottal channel. The x axis is taken as the axis of symmetry. The vocal fold motion is controlled by a two-mass model with rounded configuration [2].

pressibility, the nominal flow velocity $v_n(x)$ can be expressed as $v_n(x) = u_g / (h(x) \cdot L_g)$ where u_g is the volume velocity. However, development of the boundary layer reduces the height of the channel and hence the effective (average) flow velocity is obtained from

$$v(x) = \frac{u_g}{(h(x) - 2\delta_1(x))L_g}, \quad (1)$$

where $\delta_1(x)$ is the displacement thickness. When the glottal channel narrows, the effective velocity becomes greater than the nominal velocity. Therefore, the equation of the boundary layer needs to be solved to estimate its thickness.

2.2. Boundary-layer equation

Following the Kármán-Pohlhausen framework [3, 4, 7], the integral momentum relation of the boundary layer is formally expressed as

$$\frac{d}{dx} \{v(x)^2 \delta_2(x)\} + \delta_1(x)v(x) \frac{d}{dx} v(x) = \frac{\tau(x)}{\rho} \quad (2)$$

where $\delta_1(x)$ is the displacement thickness, $\delta_2(x)$ is the momentum thickness, $\tau(x)$ is the wall shear stress, and ρ is the air density. The core flow velocity $v(x)$ should be given to solve the integral relation. The characteristic quantities of the boundary layer, $\delta_1(x)$, $\delta_2(x)$, and $\tau(x)$, are represented as functions of the velocity in the layer.

This internal velocity is the hidden variable in the integral relation, and the additional constraint on the similarity of velocity profiles is used to solve the integral relation. We use the solution of the single parameter Falkner-Skan equation (the Hartree profile) for the family of velocity profiles [7]. When the boundary-layer equation is solved, boundary-layer thickness and wall shear stress are also determined. The core flow velocity $v(x)$ in Eq. (2), however, depends on the boundary-layer thickness, thus producing the viscous-inviscid interaction between the boundary layer and the core flow. Therefore, the equations for both the core flow and boundary layer should be solved simultaneously.

When the approximation method of Kalse *et al.* [8] is employed, the boundary-layer equation is rewritten as:

$$v(x) \frac{\delta_1(x)}{\nu} \frac{d}{dx} \frac{\delta_1(x)}{H(x)} + \left(1 + \frac{2}{H(x)}\right) f_1(H(x)) = H(x) f_2(H(x)) \quad (3)$$

and

$$\frac{\delta_1(x)^2}{\nu} \frac{dv(x)}{dx} = f_1(H(x)), \quad (4)$$

where

$$f_1(H(x)) = -2.4\{1 - \exp(0.43(2.59 - H(x)))\}, \quad (5)$$

$$f_2(H(x)) = \frac{4}{H(x)^2} - \frac{1}{H(x)}, \quad (6)$$

and $H(x) = \delta_1(x) / \delta_2(x)$.

Eqs. (1), (3) and (4) constitute a set of nonlinear simultaneous equations with respect to three unknown variables $H(x)$, $\delta_1(x)$, and $v(x)$. These variables are a function of x , and the problem can be solved numerically by using downstream marching [3]. When the interactive problem is solved, the separation point of the boundary layer, x_s , can be estimated by finding the x -axis position where the wall shear stress becomes zero. This indicates that $f_2 = 0$ and $H = 4$ at that point.

3. Volume flow estimation with coupling to the vocal-tract filter

The interactive boundary-layer analysis explained in the previous section can be solved when the volume flow or the Reynolds number is specified. Alternatively, when the subglottal pressure is specified as a phonation condition, the volume flow should be estimated. The acoustic property of the vocal tract is also represented in the form of the reflection function, and used to predict the pressure just above the glottis.

3.1. Pressure distribution in the glottis

The origin of the x -axis in Fig. 1 is taken at the stagnation point of the glottal inlet. Also, x_s , x_d , and x_r represent respectively

the position of the separation point, the glottal outlet, and the reattachment point in the tract. The pressure in the glottis can be expressed as [2]:

$$p(x) = p_B(x) + p_v(x), \quad (7)$$

where

$$p_B(x) = p_0 - \frac{1}{2}\rho \left(\frac{u_g}{L_g h(x)} \right)^2 - \frac{\rho}{L_g} \int_0^x \frac{1}{h(\tau)} d\tau \frac{du_g}{dt} \quad (8)$$

is the Bernoulli pressure and

$$p_v(x) = -\frac{12\mu}{L_g} u_g \int_0^x \frac{1}{h(\tau)^3} d\tau \quad (9)$$

is the viscous friction loss. Here p_0 is the subglottal pressure and μ is the dynamic viscous coefficient.

3.2. Reflection function of the vocal tract

Based on the theory of a lossy acoustic tube [5], the acoustic pressure P and volume velocity V at the inlet and outlet of the vocal tract can be expressed for vowel-like sounds as:

$$\begin{pmatrix} P_{out} \\ U_{out} \end{pmatrix} = K_V \times \begin{pmatrix} P_{in} \\ U_{in} \end{pmatrix}, \quad K_V = \begin{pmatrix} A_V & B_V \\ C_V & D_V \end{pmatrix}, \quad (10)$$

where $K_V = K_N \times K_{N-1} \times \dots \times K_1$ is the concatenated coefficient matrix for the entire vocal tract. Components of the matrix for the n 'th tube are expressed as $A_n = \cosh(\sigma \Delta l / c)$, $B_n = -\frac{\rho c}{S_n} \gamma \sinh(\sigma \Delta l / c)$, $C_n = -\frac{S_n}{\rho c} \frac{\sinh(\sigma \Delta l / c)}{\gamma}$, and $D_n = \cosh(\sigma \Delta l / c)$. Additionally, S_n is the sectional area, Δl the tube length, c the speed of sound, $\alpha = \sqrt{\omega c_1}$, $\beta = \frac{j\omega\omega_0^2}{(j\omega+a)j\omega+b} + \alpha$, $\gamma = \sqrt{\frac{\alpha+j\omega}{\beta+j\omega}}$, and $\sigma = \gamma(\beta + j\omega)$. Values for a , b , c_1 , and ω_0 were obtained from the literature [5].

The reflection function $r_V(t)$ is defined by the inverse Fourier transform of $R_V = (Z_V - Z_0)/(Z_V + Z_0)$, where $Z_V = (D_V Z_L - B_V)/(A_V - C_V Z_L)$ is the input impedance of the vocal tract, $Z_0 = \rho c / S_1$ is the characteristic impedance of the first section, and Z_L is the radiation impedance of the lips. We assume here that Z_V gives the ratio of the pressure at the flow reattachment point to the glottal flow, $Z_V = P_r / U_g$, where P_r and U_g are Fourier transforms of the temporal change of $p_r = p(x_r)$ and u_g , respectively.

The acoustic pressure at the inlet of the vocal tract can then be obtained as [5]:

$$p_r(t) - r_V(t) * p_r(t) = Z_0 u_g(t) + Z_0 r_V(t) * u_g(t), \quad (11)$$

where $*$ denotes convolution. This leads to the following discrete-time representation:

$$p_r(n) = \frac{1 + r_V(0)}{1 - r_V(0)} Z_0 u_g(n) + \frac{1}{1 - r_V(0)} \sum_{k=1}^{N-1} r_V(k) \left\{ p_r(n-k) + Z_0 u_g(n-k) \right\}, \quad (12)$$

where N is the length of the reflection function.

3.3. Estimation of the volume flow

We show next how the volume flow can be determined. Since the boundary-layer analysis is effective for the region upstream of the separation point, another physical constraint is required to incorporate the pressure downstream of the glottis. In this study,

the conventional flow separation-reattachment framework [2, 6] is employed for this problem. Momentum conservation upstream and downstream of the separation point implies that [2, 3, 6]:

$$p(x_d) S(x_r) + \frac{\rho u_g^2}{S(x_d)} = p(x_r) S(x_r) + \frac{\rho u_g^2}{S(x_r)}, \quad (13)$$

where $S(x)$ is the sectional area. If we assume that the sectional area of the glottal jet is constant within the glottis, the pressure and jet width is constant between the separation point and the glottal exit such that $p(x_d) = p(x_s)$ and $S(x_d) = S(x_s)$ [2], giving:

$$p(x_s) S(x_r) + \frac{\rho u_g^2}{S(x_s)} = p(x_r) S(x_r) + \frac{\rho u_g^2}{S(x_r)}. \quad (14)$$

In this equation, $p(x_s)$ is given by Eq. (7) and $p(x_r)$ by Eq. (12). $S(x_s) = L_g (h(x_s) - 2\delta_1(x_s))$ is the effective sectional area determined from the solution of the boundary-layer analysis. $S(x_r)$ may be set as $S(x_r) = S_1$, the area for the first vocal-tract section. The volume flow u_g can then be determined as:

$$u_g(n) = \frac{-A + \sqrt{A^2 + 4M(p_0 + \alpha u_g(n-1) - d)}}{2M} \quad (15)$$

where $A = a + b + c$, $a = \rho / (L_g \Delta t) \int_0^{x_s} 1/h(x) dx$, $b = 12\mu / L_g \int_0^{x_s} 1/h(x)^3 dx$, $c = Z_0 \{1 + r_V(0)\} / \{1 - r_V(0)\}$, $d = 1 / \{1 - r_V(0)\} \sum_{k=1}^{N-1} r_V(k) \left\{ p_r(n-k) + Z_0 u_g(n-k) \right\}$, $M = \rho(1 - 2N + 2N^2) / \{2A(x_s)^2\}$, $N = A(x_s) / A(x_r)$, and Δt is the time difference.

3.4. Mechanical model of the vocal fold

The motion of the vocal fold is represented by a model similar to the two-mass model [6]. To form a smooth glottal channel, each vocal fold is shaped by using two circles and three line segments connecting the inlet and outlet of the glottis and the circles (Fig. 1). Masses m_1 and m_2 are assigned to each circle, and are connected to the fixed wall by dampers of resistances r_1 and r_2 and linear springs with Hooke constants k_1 and k_2 . The two masses are also connected by a linear spring of constant k_{12} . The equations of vocal fold motion are given as [2]:

$$m_1 \frac{d^2 y_1}{dt^2} + r_1 \frac{dy_1}{dt} + s_1 + k_{12}(y_1 - y_2) = F_1, \quad (16)$$

$$m_2 \frac{d^2 y_2}{dt^2} + r_2 \frac{dy_2}{dt} + s_2 + k_{12}(y_2 - y_1) = F_2, \quad (17)$$

and

$$s_i = \begin{cases} k_i y_i & (y_i + y_0 > 0) \\ 4k_i y_i & (y_i + y_0 \leq 0) \end{cases}, \quad (18)$$

where y_i is the displacement of each mass and y_0 is the rest position. We assume that the flow pressure as expressed by Eq. (7) acts on the lower mass through F_1 [2], while the acoustic pressure of Eq. (12) acts on both masses.

4. Simulation results

Voice production simulations were performed using the following steps. Given the subglottal pressure p_0 and the initial mass displacement y_0 , the channel height $h(x)$ was determined. The interactive boundary-layer analysis and estimation of the volume flow were then performed [3], providing information on

the volume flow u_g and pressure distribution $p(x)$. The pressure p_r at the vocal tract inlet was then calculated from the time history of u_g and p_r , and $p(x)$ and p_r were used to determine the driving force applied to the vocal folds. The mechanical equations were solved using the Runge-Kutta method, and the displacement of each mass was updated. These steps were repeated for specified time durations.

The subglottal pressure p_0 was set to 8 cmH₂O. The vocal fold length L_g was 1.2 cm. Parameters of the vocal fold model were set as $m_1 = 0.125/Q$ [g], $m_2 = 0.025/Q$ [g], $k_1 = 80000Q$ [dyn/cm], $k_2 = 8000Q$ [dyn/cm], $k_{12} = 25000Q$ [dyn/cm], $r_1 = 0.2\sqrt{m_1 k_1}$ [dyn-sec/cm], and $r_2 = 1.2\sqrt{m_2 k_2}$ [dyn-sec/cm], respectively [2, 6]. Q was the tension parameter. The initial displacement y_0 was 0.0025 [cm], and the sampling frequency was 20 kHz.

Figure 2 shows results of preliminary simulations for vowels /a/ and /i/, where the tension parameter was set as $Q = 1$ or 3. Each trace shows the temporal waveform for the volume flow u_g , pressure p_r , mass displacements y_1 and y_2 , and the synthetic speech obtained by convolving the volume flow waveform with the inverse Fourier transform of the vocal tract transfer function. The displacements y_1 and y_2 are shown by the solid and broken lines, respectively. The frequency for the first formant F_1 was about 800 Hz for /a/ and 400 Hz for /i/ in these simulations. Because the fundamental frequency F_0 of the vocal fold vibration was about 154 Hz, F_0 was lower than F_1 for the /a/ simulations, while it was higher than F_1 for /i/.

Simulated variations of the variables due to the Q parameter settings seem to be acceptable in general. In the /a/ simulations, the value of p_r becomes negative in accordance with the closing phase of the vocal fold motion, thus enhancing their oscillation. The amplitude of p_r waveform was higher for /i/ than for /a/, implying that the vocal fold motion and volume flow were more strongly affected by p_r for /i/. Although the fundamental frequency was higher than the F_1 for /i/, unfavorable source-filter interaction [1] was not observed explicitly, possibly due to the absence of the sub-glottic acoustic cavity in the present study.

5. Conclusions

A model was presented to explain the self-oscillation of vocal folds and the periodic changes of glottal volume flows. The model was constructed with an accurate flow analysis method based on the boundary-layer approximation, with which the effective flow channel was obtained by considering the thickness of the layer. In addition, the variable flow separation point was used to estimate the volume flow. The simulation results indicate that the model is capable of predicting the essential aerodynamic and acoustic phenomena in human phonation, but addition of the sub-glottic acoustic cavity is clearly needed to examine the source-filter interaction [1]. Research partly supported by the Grant-in-Aid for Scientific Research from the JSPS (Grant No. 18500134 and 19103003).

6. References

- [1] Titze, I. R., "Nonlinear source-filter coupling in phonation: Theory", *J. Acoust. Soc. Am.*, 123(5):2733–2749, 2008.
- [2] Pelorson, X., Hirschberg, A., van Hassel, R. R., Wijnands, A. P. J., and Auregan, Y., "Theoretical and experimental study of quasisteady-flow separation within the glottis during phonation. Application to a modified two-mass model", *J. Acoust. Soc. Am.*, 96(6):3416–3431, 1994.
- [3] Kaburagi, T., "On the viscous-inviscid interaction of the flow

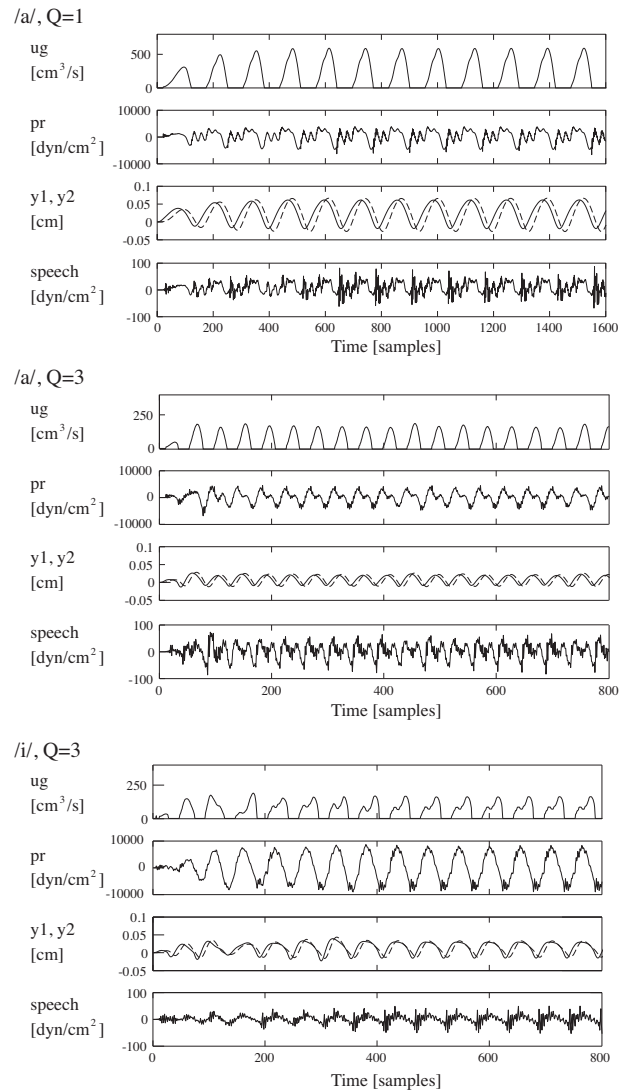


Figure 2: Results of voice production simulations.

passing through the glottis", *Acoust. Sci. & Tech.*, 29(2):167–175, 2008.

- [4] Kaburagi, T. and Tanabe, Y., "Low-dimensional models of the glottal flow incorporating viscous-inviscid interaction", *J. Acoust. Soc. Am.*, 125(1):391–404, 2009.
- [5] Sondhi, M. M., and Schroeter, J., "A hybrid time-frequency domain articulatory speech synthesizer", *IEEE Trans. Acoust., Speech & Signal Process.*, ASSP-35(7):955–967, 1987.
- [6] Ishizaka, K., and Flanagan, J. L., "Synthesis of voiced sounds from a two-mass model of the vocal cords", *Bell Syst. Tech. J.*, 51(6):1233–1268, 1972.
- [7] Schlichting, H., *Boundary-Layer Theory*, 8th Ed., Springer Verlag, 1999.
- [8] Kalse, S. G. C., Bijl, H., and van Oudheusden, B. W., "A one-dimensional viscous-inviscid strong interaction model for flow in indented channels with separation and reattachmen", *ASME J. Biomech. Eng.*, 125(3):355–362, 2003.