Frequency-domain representation of source-filter coupling and its effect in the production of voice

Tokihiko Kaburagi [1]

[1] Faculty of Design, Kyushu University
4-9-1 Shiobaru, Minami-ku, Fukuoka 815-8540 Japan.
kabu@design.kyushu-u.ac.jp

10.21437/Interspeech.2011-687

Abstract

The acoustic coupling between the voice production system and the vocal tract has a significant influence on the production of voice. In this study, the coupling effect was represented using the acoustic pressure difference across the glottis, which is capable of inducing a flow, and the mean acoustic pressure in the glottis, which acts as a driving force for the vocal folds. These specific acoustic pressures were then interpreted in the frequency domain in the form of frequency responses, and incorporated into a model of the voice production system. In this framework, we were able to test the effect of source-filter coupling by filtering frequency responses. Numerical results revealed that these responses and the input impedance of the vocal tract both exhibited a dominant peak around 4 kHz. In addition, voice production simulations revealed that this high-frequency peak has a significant influence on the spatio-temporal pattern of glottal volume flow and vocal fold movements.

Index Terms: voice production, source-filter coupling, vocal fold, glottal flow

1. Introduction

In the production of voiced speech, glottal sound sources generated by oscillations of the vocal folds travel through the vocal tract and form a sound field in the vicinity of the glottis. This acoustic feedback from the tract influences the flow though the glottis and movements of the vocal folds, a phenomenon known as source-filter coupling. This coupling has various nonlinear effects, such as flow skewing [1], enhancement or suppression of vocal-fold oscillations [2, 3], lowering oscillation threshold pressure [4], and discontinuity of changes in the fundamental frequency caused by a continuous changes in vocal fold parameters [5, 6].

Time-domain computer simulation has been an important tool for investigating these 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 [5]. It has been shown that the effect becomes more intense when the fundamental frequency approaches the resonant frequency. In addition, vocal fold oscillation becomes unstable when the fundamental frequency is between the resonant and antiresonant frequencies. Titze [3] interpreted this effect in terms of the reactance of impedances arising from the acoustic load of the sub- and supra-glottal tracts. It was shown that the harmonic component of the glottal volume flow waveform becomes weak when the reactance is negative at that frequency (level one interaction). It was also explained that the reactance of vocal-tract impedance should be inertive, while that of the sub-glottal tract should be compliant for the oscillation of the vocal folds (level two interaction).

In the current paper, source-filter coupling is represented in the frequency-domain, as in Titze's previous study [3]. The acoustic tubes of the sub- and supra-glottal tracts are connected to the intermediate part of the glottis, and the frequency response of the acoustic pressures near the glottis is expressed relative to each acoustic component of the glottal volume flow. On the basis of this acoustic impedance representation, we examined the effects of the acoustic pressure difference across the glottis and the mean intra-glottal pressure. In addition, these frequency responses were low-pass filtered, and the consequences of this frequency-domain manipulation of the source-filter coupling were investigated using a time-domain simulation of the voice production process.

2. Representation of source-filter coupling

2.1. Specific acoustic pressures related to the coupling effect

Source-filter coupling constitutes an acoustic influence of the pressure near the glottis on the glottal flow and movements of the vocal folds. Two specific pressure values are used here to represent these effects. One of these values is the difference between the inlet pressure (p_{A0}) and outlet pressure (p_{A1}) of the glottis

$$\Delta p_A = p_{A0} - p_{A1} \ , \tag{1}$$

which causes the pressure gradient across the glottis, and is used to estimate the volume flow together with the static lung pressure [Eqs. (5) and (6)]. The other is the mean acoustical pressure in the glottis

$$\bar{p}_A = \frac{p_{A0} + p_{A1}}{2} \ , \tag{2}$$

which can be used to estimate the driving force applied to the vocal folds in addition to the aerodynamic pressure of the glottal flow [Eqs. (11) and (12)]. To discriminate between these temporal waveforms and their Fourier transforms, time-domain variables are denoted by lower-case letters while frequency-domain variables are denoted by capital letters in the following expressions.

The values of these specific acoustic pressures are determined on the basis of the model shown in Fig. 1. From this model, relationships $P_{A0} = -Z_0 U_{A0}$, $P_{A1} = Z_1 U'_{A1}$, and $U'_{A1} = U_{A1} + U_g$ are obtained, where Z_0 and Z_1 are the input impedance of the sub- and supra-glottal tracts. The input impedance of the glottis seen from the vocal tract, Z_g, can be expressed as $Z_g = -P_{A1}/U_{A1} =$

Copyright © 2011 ISCA

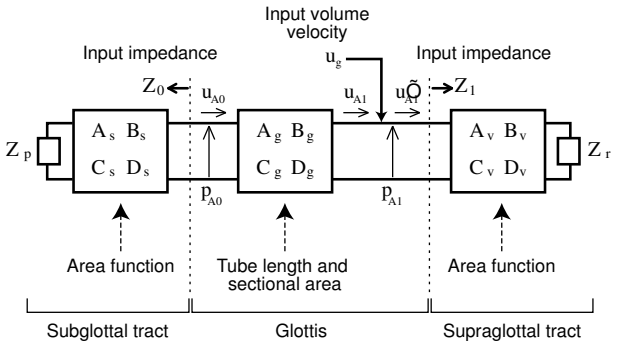


Figure 1: *Acoustic tube model connecting the sub- and supra-glottal tracts via the glottis.*

$-(B_g - A_g Z_0)/(D_g - C_g Z_0)$, where A_g, B_g, C_g, and D_g are elements of the transmission matrix of the glottis approximated as a tube with an uniform cross section.

From these relationships, ΔP_A and $\bar{P}_A$ are each obtained as the frequency response of a filter having the acoustic component of the glottal volume flow as its input:

$$Z_\Delta = \frac{\Delta P_A}{U_g} = \frac{P_{A0} - P_{A1}}{U_g} = \frac{\{B_g - (A_g - 1)Z_0\}Z_1}{Z_D} \quad (3)$$

and

$$Z_M = \frac{\bar{P}_A}{U_g} = \frac{P_{A0} + P_{A1}}{2U_g} = -\frac{\{B_g - (A_g + 1)Z_0\}Z_1}{2Z_D} , \quad (4)$$

where $Z_D = (D_g - C_g Z_0)Z_1 - (B_g - A_g Z_0)$. The input impedances of the tracts are obtained as $Z_0 = -(A_s Z_p + B_s)/(C_s Z_p + D_s)$ and $Z_1 = (D_v Z_r - B_v)/(A_v - C_v Z_r)$, where A_s, B_s, C_s, and D_s are elements of the transmission matrix for the sub-glottal tract and A_v, B_v, C_v, and D_v those for the supra-glottal tract. The values of these elements are calculated by applying an acoustic tube model of the lossy vocal tract [7], taking into account data regarding vocal-tract area function [8]. Z_p and Z_r are the terminal impedance for the lungs and lips, respectively.

2.2. Total framework of the voice production model

Figure 2 shows the framework of the voice production model to examine the influence of the acoustic feedback of the tract on the process of the voice production. The frequency responses, Z_Δ and Z_M, are determined from the area data as explained in the preceding subsection. In some cases, these must be processed by a zero-phase low-pass filter to change the coupling effect. Next, the inverse Fourier transforms of the frequency responses, z_Δ and z_M, are convolved with the waveform of glottal volume flow, u_g, and the values of the specific acoustic pressures, Δp_A and $\bar{p}_A$, are calculated.

The voice production model is composed of a method for estimating the glottal volume flow and a model of the vocal folds. 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 [9]. Here, the pressure difference across the glottis is determined by summing the static sub-glottal pressure supplied by the lungs, p_{F0}, and the acoustic pressure as

$$\Delta p = p_{F0} + \Delta p_A = p_{F0} + z_\Delta * u_g . \quad (5)$$

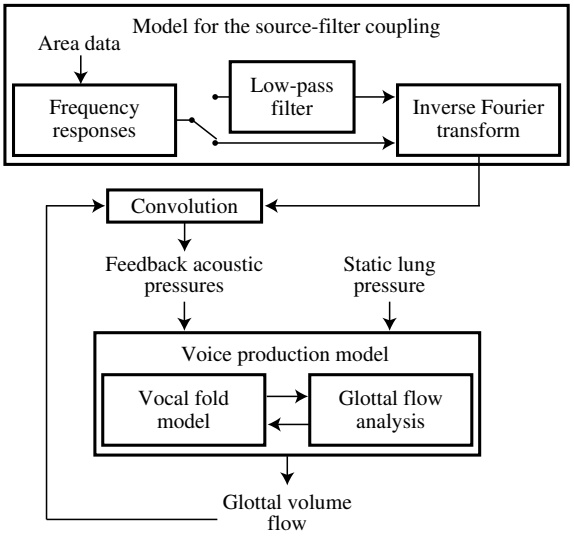


Figure 2: *Framework of the voice production model incorporating the effect of source-filter coupling.*

The convolution is computed only for the latest time instant. From the Bernoulli relation, we obtain a pressure-flow relation

$$\Delta p = 0.5\rho \frac{u_g^2}{S_s^2} . \quad (6)$$

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

$$S_s = \{h(x_s) - 2\delta(x_s)\}l_g . \quad (7)$$

l_g is the length of the vocal fold, $h(x)$ is the channel height, $\delta(x)$ is the displacement thickness of the boundary layer, and x_s is the point of flow separation along the x axis. The x axis is taken as the line of symmetry between both vocal folds. ρ is the density of the air. From these relationships, the volume flow is estimated in the discrete-time form as

$$u_g(n) = \frac{z_\Delta(0)S_s^2 + S_s\sqrt{(z_\Delta(0)S_s)^2 + 2\rho P}}{\rho} \quad (8)$$

where $P = p_{F0} + \sum_{k=1}^{K-1} z_\Delta(k)u_g(n-k)$ and n represents the time index. The flow separation point, x_s, and the displacement thickness, $\delta(x)$, are obtained from the boundary-layer analysis [9]. If the value of the volume flow is estimated at an instant, the effective flow velocity is determined as $v(x) = \frac{u_g}{\{h(x) - 2\delta(x)\}l_g}$ and the pressure distribution is determined as $p_F(x) = p_{F0} - \frac{1}{2}\rho v(x)^2$, allowing the driving force of the vocal folds to be determined.

Movement of the vocal folds is modeled using the modified two-mass model [10]. 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. 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 motion is given as

$$m_1 \frac{d^2 y_1}{dt^2} + r_1 \frac{dy_1}{dt} + k_1 y_1 + k_{12}(y_1 - y_2) = f_1 \quad (9)$$

and

$$m_2 \frac{d^2 y_2}{dt^2} + r_2 \frac{dy_2}{dt} + k_2 y_2 + k_{12}(y_2 - y_1) = f_2 \ , \qquad (10)$$

where 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 driving force is exerted only on the lower, first mass, and f_2 is set to zero. The driving force, f_1, is determined as follows. The entire driving pressure is given by summing the aerodynamic and acoustic pressures as

$$p(x) = p_F(x) + \bar{p}_A = p_{F0} - \frac{1}{2}\rho v(x)^2 + z_M * u_g \ , \qquad (11)$$

where the convolution is computed for the latest time instant. The driving force for the lower mass is then given as

$$f_1 = \lambda l_g \left\{ \int_{x_0}^{x_s} p_F(x) dx + (x_e - x_0)\bar{p}_A \right\} \ . \qquad (12)$$

Here, x_0 is the inlet of the glottis, x_e is the outlet of the glottis, 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 $f_1 = \lambda l_g (x_e - x_0) p_{F0}$.

3. Numerical results

3.1. Frequency responses representing the coupling effect

The frequency responses of the specific acoustic pressures, Z_Δ and Z_M, were computed from data related to vocal-tract area function taken for the vowel /I/ in "hid" [8] and data related to the sub-glottal tract [11]. The length of the glottal tube (depth of the glottis) was set to 0.3 cm, and the cross-sectional area was set to 0.014 cm^2. The results are plotted in Fig. 3.

From the top plot, it can be seen that the input impedance of the vocal tract, Z_1, exhibited a significant peak at approximately 4.1 kHz. This peak is much stronger than those corresponding to the first to fourth formants. An investigation confirmed that this high-frequency peak was due to the resonance of epiglottal tube, i.e., a local, small acoustic cavity above the glottis. The figure also shows that each peak in the input impedance forms a peak of the same frequency in the magnitude spectrum of the responses, Z_Δ and Z_M, plotted in Fig. 3 (B). Concerning the peaks seen in the input impedance, Titze [3] clarified that source-filter coupling can be enhanced by reducing the cross-sectional area of the epiglottal tube. It should also be noted that the magnitude of frequency responses is interpreted as the frequency-dependent strength of source-filter coupling, because they determine the amplitude of the pressure values, Δp_A and $\bar{p}_A$, resulting from the volume flow, u_g.

The influence of the high-frequency peak at 4.1 kHz is examined through a time-domain simulation, below. Here, as shown in Fig. 2, the frequency responses (Z_Δ and Z_M) are low-pass filtered with a cut-off frequency of 3.8 kHz at each time step of the simulation. The bottom plot of Fig. 3 shows the filtered versions of the responses shown in Fig. 3 (B).

3.2. Dynamic simulation of voice production

In the simulation, the parameters of the vocal-fold model were set to $m_1 = 0.125$ g, $m_2 = 0.025$ g, $k_1 = 80000$ dyn/cm, $k_2 = 8000$ dyn/cm, $k_{12} = 55000$ 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. For the glottis in a closed state, the values of the stiffness parameters were increased such that $k_1 = 320000$ and $k_2 = 32000$ dyn/cm. The dampers were also increased such that $r_1 = 257$ and $r_2 = 49.6$ dyn·s/cm. The initial displacement was $y_0 = 0.00014$ cm and the vocal-fold length was $l_g = 1.2$ cm. λ was determined so that the effective depth of the lower mass, $\lambda(x_e - x_0)$, was approximately 0.27 cm. The sub-glottal pressure was 8 cmH_2O and the sampling frequency was 20 kHz.

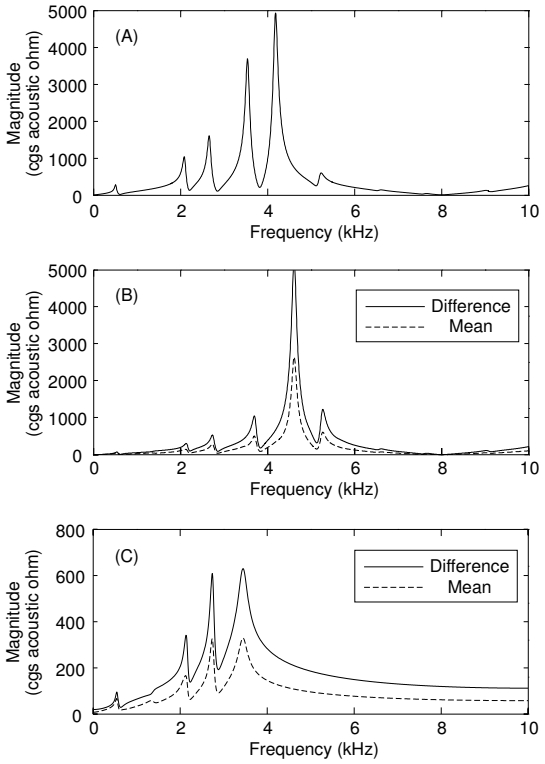


Figure 3: *Computed results of (A) vocal-tract input impedance, (B) frequency responses of the acoustic pressure difference Z_Δ (solid) and mean acoustic pressure Z_M (broken) for the vowel /I/, and (C) the filtered versions of the responses.*

The plots in Fig. 4 show the changes with time of the glottal volume flow (u_g), effective glottal area at the flow separation point (S_s), acoustic pressure difference across the glottis (Δp_A), mean acoustic pressure in the glottis ($\bar{p}_A$), mass displacements (y_1 and y_2), and the driving force on the vocal folds (f_1). The broken line in the driving force plot shows the force component generated by the aerodynamic pressure, while the solid line is the sum of aerodynamic and acoustic pressures. The volume flow is basically in proportion to the glottal area, but the acoustic pressure difference in this simulation was positive when the glottis was closing. This had the effect of inducing a forward flow and sharpening the waveform of the glottal flow. On the other hand, the aerodynamic and acoustic forces were both negative when the glottis was closing. The glottis was then effectively closed and the vocal fold oscillation was maintained.

Finally, Fig. 5 shows the results of a simulation where the frequency responses, Z_Δ and Z_M, were low-pass filtered (see the bottom plot of Fig. 3) to eliminate peaks higher than 4kHz. In the figure, the acoustic pressure difference was almost negative, which produces a backward flow, and reduces the volume flow through the glottis. On the other hand, the amplitude of

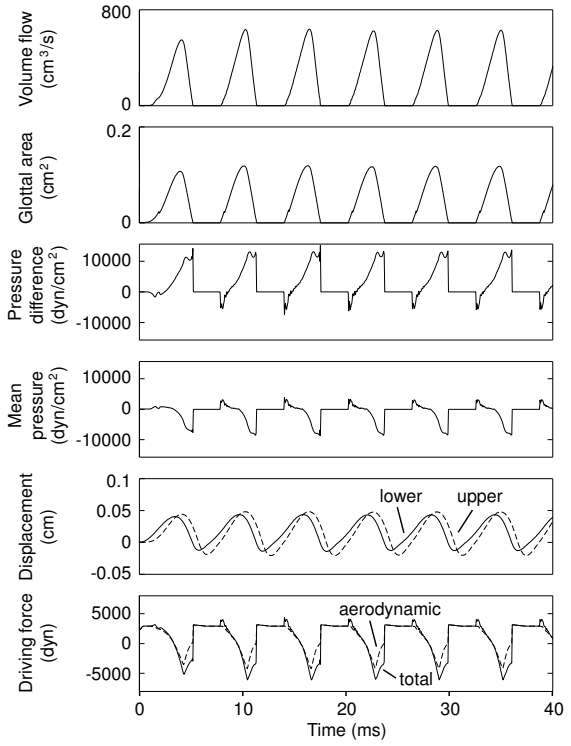


Figure 4: *Results of the voice production simulation for the vowel /I/.*

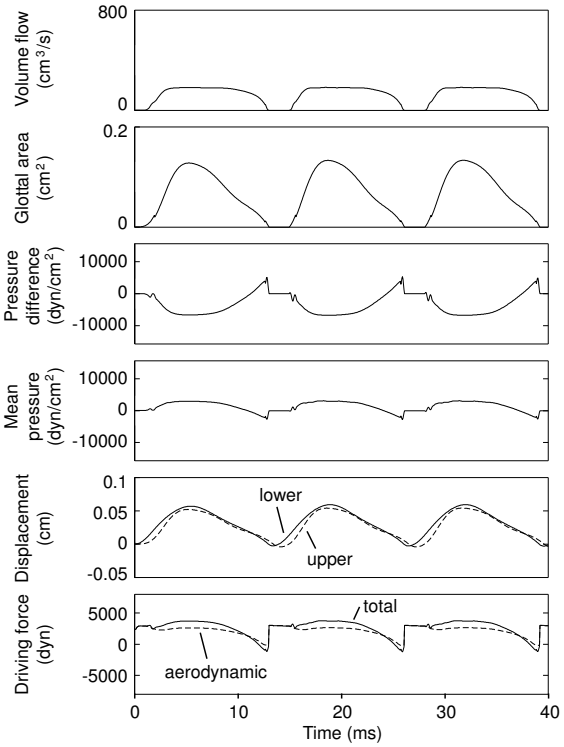


Figure 5: *Results of the voice production simulation with the filtering of the specific frequency responses, Z_Δ and Z_M.*

the mean acoustic pressure was very small, and the total driving force was slightly negative when the glottis was closing. A comparison of these results with those shown in Fig. 4 reveals that the waveforms of the glottal area and vocal fold displacement were differently shaped, with the period of vocal fold oscillation almost doubling. The movements of the lower and upper masses were almost synchronized, indicating that convergent-divergent changes in the glottal channel were not exhibited. As such, the elimination of the high-frequency component of Z_Δ and Z_M caused a significant change in the motion of the vocal folds and the waveform of the glottal volume flow.

4. Conclusion

Source-filter coupling is typically observed when the frequency of the vocal fold oscillation approaches the resonance frequency of the vocal tract [5]. Titze [3] also reported that each frequency component in the glottal volume flow interacts with the vocal-tract resonator independently. In the current study, the effect of the source-filter coupling was investigated further using the frequency response of two types of acoustic pressures. The results of voice production simulations clearly indicated that the high-frequency component of these frequency responses has a strong effect on the processes involved in voice production.

This research was partly supported by the Grant-in-Aid for Scientific Research from the Japan Society for the Promotion of Science (Grant No. 19103003).

5. References

[1] Rothenberg, M., "Acoustic interaction between the glottal source and the vocal tract", in Stevens and Hirano [Eds], Vocal Fold Physiology, 305–328, University of Tokyo Press, 1981.

[2] Zañartu, M., Mongeau, L., and Wodicka, G. R., "Influence of acoustic loading on an effective single mass model of the vocal folds", J. Acoust. Soc. Am., 121(2): 1119–1129, 2007.

[3] Titze, I. R., "Nonlinear source-filter coupling in phonation: Theory", J. Acoust. Soc. Am., 123(5): 2733–2749, 2008.

[4] Chan, R. W. and Titze, I. R., "Dependence of phonation threshold pressure on vocal tract acoustics and vocal fold tissue mechanics", J. Acoust. Soc. Am., 119(4): 2351–2362, 2006.

[5] 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.

[6] Švec, J. G., Schutte, H. K., and Miller, D. G., "On pitch jumps between chest and falsetto registers in voice: Data from living and excised human larynges", J. Acoust. Soc. Am., 106(3): 1523–1531, 1999.

[7] Sondhi, M. M. and Schroeter, J., "A hybrid time-frequency domain articulatory speech synthesizer", IEEE Trans. Acoust., Speech & Signal Process., ASSP-35: 955-967, 1987.

[8] Story, B. H., Titze, I. R., and Hoffman, E. A., "Vocal tract area functions from magnetic resonance imaging", J. Acoust. Soc. Am., 100(1): 537–554, 1996.

[9] 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.

[10] 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.

[11] Weibel, E. R., Morphometry of the human lung, 136–140, Springer, 1963.