HIGH-PITCHED VOICE SIMULATION USING A TWO-DIMENSIONAL VOCAL FOLD MODEL

Seiji Adachi† and Jason Yu‡
ATR Human Information Science Laboratories, Keihanna Science City, Kyoto 619-0288 Japan†
School of Engineering Science, Simon Fraser University, 8888 University Dr. Burnaby, BC V5A 1S6 Canada‡

Voiced sounds were simulated with a computer model of the vocal fold composed of a single mass vibrating both parallel and perpendicular to the airflow. The major improvement of the present model over the two-mass model is that it has a wider continuous frequency range where self-excitation is possible both below and above the first formant frequency of the vocal tract. The two-dimensional model can therefore successfully be applied to the sound synthesis of a high-pitched soprano singing, where the fundamental frequency sometimes exceeds the first formant frequency.

I. Introduction

An acoustic tube generally yields an inductive load in the frequency below a resonance peak, while the load turns out to be capacitive in the frequency above the peak. In speech, the fundamental frequency (F0) is usually lower than the first formant frequency (F1) of the vocal tract. Therefore, the vocal folds always vibrate with an inductive load.

In high-pitched soprano singing, however, F0 enters the range of F1 in normal speech. The soprano singer then raises F1 as F0 approaches F1 by increasing the jaw opening.[1] As a result, F1 is always tuned close to F0. A more recent measurement of the vocal tract resonance[2, 3] confirms this effect in the middle range of soprano singing. It also shows that F1 cannot be raised above a certain point (roughly 1 kHz), and the order of F0 and F1 is reversed in the high range. These observations imply that the vocal folds should vibrate in the vicinity of the frequency region near F1 where the acoustic load can be both inductive and capacitive.

In speech synthesis by simulating all the processes in the voice production, the two-mass model of the vocal fold vibration[4] has been widely used. It was shown that this model can simulate self-sustained oscillation with a capacitive acoustic load of the vocal tract. As shown in Section IV, however, self-sustained oscillation can not be obtained in a large frequency range just above F1. This means that musical tones in this range can not be synthesized. The voice range simulated by the two-mass model, therefore, becomes narrower.

†sadachi@atr.jp

This paper proposes a model of vocal fold vibration that can successfully simulate self-excited oscillation in a wide frequency range on both sides of F1 with no discontinuity of vibration. The model approximates the vocal folds as a pair of single masses that can vibrate both parallel and perpendicular to the airflow. The high-pitched singing voice is simulated as well as the normal speech voice in this paper. In addition, the mechanism for self-excitation both in the capacitive and inductive acoustic loads is discussed.

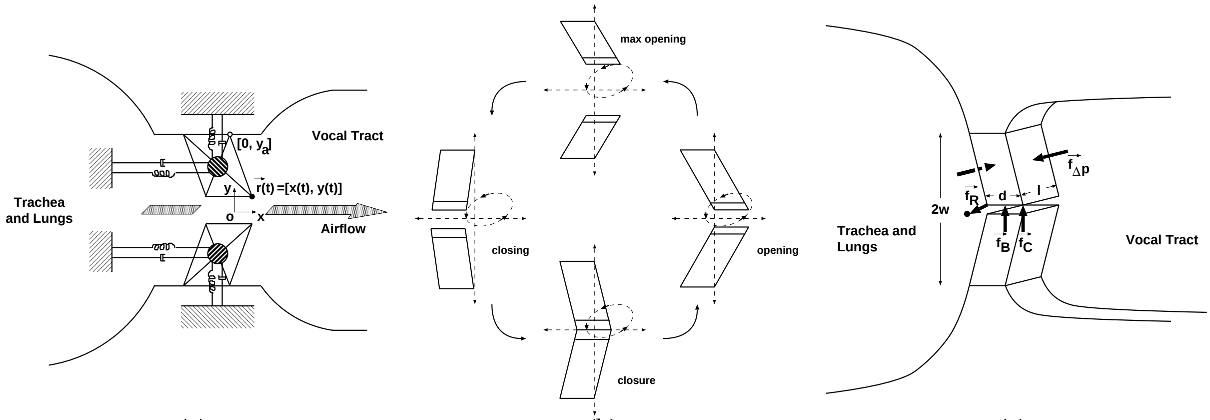
II. Two-dimensional model of vocal folds

The proposed model was originally developed to simulate the vibration of a brass player's lips.[5] The model is a combination of two earlier models: the swinging-door model and the transverse model.[6] The former employs a valve (lips or vocal folds) that operates by the pressure difference between the upstream and downstream regions. The latter employs a valve that operates by the Bernoulli pressure generated by a flow passing through the valve aperture.

The complete description of the model including the equation of motion can be found in [7]. The pair of modeled vocal folds are schematically represented by parallelograms with two sets of a spring and a damper as shown in Fig. 1 (a). The left and right vocal folds are assumed to vibrate symmetrically. The vocal fold simultaneously executes both swinging and elastic motions. The former is driven by the trans-glottal pressure difference, and the latter is driven by the Bernoulli pressure generated at the glottis. The swinging motion implies that the motions parallel and perpendicular to the airflow are not independent but coupled with each other. This differentiates the current model from other two-dimensional models, such as those proposed by Liljencrants[8] and Flanagan and Ishizaka[9]. During one cycle of oscillation depicted in Fig. 1 (b), the tip of the vocal fold makes an elliptic trajectory, while the glottis retains a rectangular shape.

Forces acting on the vocal fold are illustrated in Fig. 1 (c). These are Bernoulli force $\vec{f}_{\rm B}(t)$ in the glottis, force due to the pressure difference $\vec{f}_{\Delta p}(t)$, contact force $\vec{f}_{\rm C}(t)$, and restoring force $\vec{f}_{\rm R}(t)$ from both springs. Figure 1 (c) also shows the lateral dimension (width) w and the length

Models and analysis of vocal emissions for biomedical applications. 4th international workshop. October 29-31, 2005 – Firenze, Italy.
Edited by C. Manfredi. ISBN 88-8453-320-1 (online)



(a) (b) (c)

Figure 1: The two-dimensional model of vocal fold vibration: (a) Schematic diagram of the model. (b) Motion of the vocal folds allowed in this model. Four different phases in a single cycle of oscillation are shown. (c) Forces acting on the vocal fold and the dimensions of the vocal fold.

l and thickness d of the vocal fold. To simulate vocal fold vibration, we have to know the external forces $\vec{f}_{\mathrm{B}}(t)$ and $\vec{f}_{\Delta p}(t)$, both acting on the vocal folds. These forces are determined by the acoustic response of the vocal tract and fluid dynamics.

Synthesized sound can be obtained by solving the equation of motion for the vocal fold vibration, a modeled equation for the air flowing through the glottis and the input impedance of the vocal tract that can be determined from the shape of the vocal tract by the transmission line method. The parameters that can control the system of the voice production are the rest position of the vocal fold (x_0, y_0), subglottal pressure p_0 and the resonance frequency of the vocal fold f_r.

III. Vowel simulation

As a typical sound generated by the two-dimensional vocal fold model, simulation result of vowel /e/ is shown in Fig. 2. The control parameters are set to $x_0 = 0.2$ mm, $y_0 = -0.02$ mm, $p_0 = 800$ Pa and $f_r = 120$ Hz.

The trajectory of the vocal fold has a smooth oval shape, and the oscillation is in the counter-clockwise direction for the upper mass. This result is in accord with common observations of vocal fold vibration.[10] The glottal area waveform has a symmetric shape in the opening and closing phases with the duty ratio of 0.68. The simulated glottal flow is a more symmetric shape than that assumed by the Rosenberg model[11]. This may be due to the inclusion of the flow generated by the mechanical motion of the vocal folds. The simulated pressure waveforms both at the entrance and at the exit of the vocal tract resemble those by the two-mass model.

The five Japanese vowels /i/, /e/, /a/, /o/ and /u/ are simulated with the vocal tract shapes obtained from a Japanese male speaker. The sound spectra of the simulated vowels are shown in Fig. 3. F0's of the simulated sounds are 126.3, 133.7, 134.1, 130.8 and 127.7 Hz for vowels /i/, /e/, /a/, /o/ and /u/, respectively. Each simulated vowel has the typical first and second formant frequencies. A simple listening test also confirms that the simulated vowels have the same quality as those synthesized with the two-mass model.

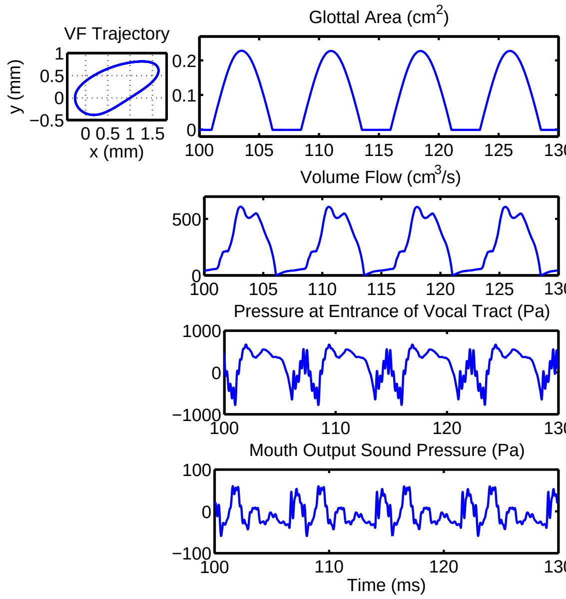


Figure 2: Simulation results for vowel /e/ showing trajectory of the vocal fold, glottal area, volume flow, pressure at entrance of vocal tract, and output pressure.

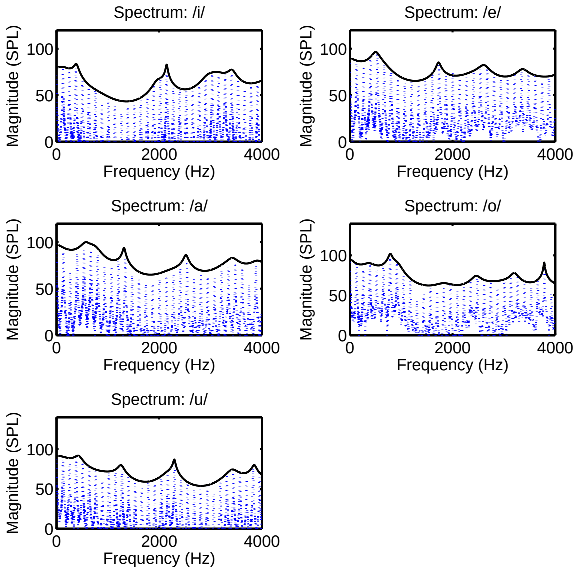


Figure 3: Simulated sound spectra for vowels /i/, /e/, /a/, /o/ and /u/. The solid lines show the spectrum envelopes estimated by LPC analysis.

IV. High-pitched voice simulation

The response of the two-dimensional and two-mass models with a capacitive acoustical load was investigated by driving the oscillation frequency to values between the first resonance frequency (F1) and the first anti-resonance frequency (F1′).

A. Straight tube load

The two vocal fold models were attached to a cylinder of 17 cm length and 5 cm^2 cross-sectional area. Calculated F1 and F1′ are 516 and 977 Hz, respectively, the capacitive region lies between F1 and F1′. The models were then driven at the range of the vocal fold resonance frequency f_r, and the sound frequency was measured. When increasing the f_r, the subglottal pressure p_0 should also be increased to obtain self-excitation. The other parameters x_0 and y_0 are set to 0.2 and -0.05 mm, respectively, in this experiment. The relationship between sound frequency and f_r for the two-dimensional model is plotted with a solid line in Fig. 4 (a). The relationship for the two-mass model is shown in Fig. 4 (b). In these figures, p_0 change is also plotted with dash-dot lines.

When the sound frequency increases beyond F1, the acoustic load changes from the inductive to capacitive behavior. The sound frequency of the two-dimensional model increases smoothly with f_r. Self-excitation is possible in the capacitive region continuously nearly up to F1′. On the other hand, the two-mass model has a jump in frequency at

(a)

(b)

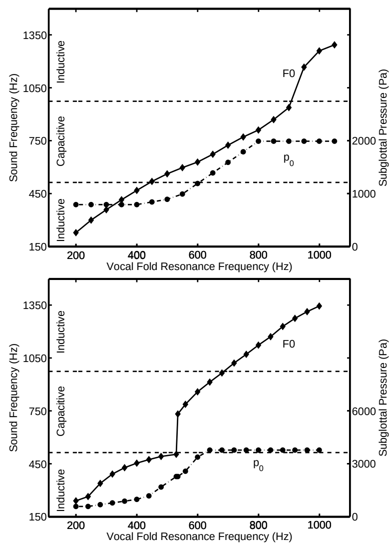


Figure 4: Straight-tube simulation. (a) Two-dimensional model. (b) Two-mass model.

F1. The jump covers about the lower half of the capacitive region. Therefore, no self-excitation can be generated between 506 and 735 Hz. Hence, the two-dimensional model is capable of producing a wider continuous range of frequencies for self-excitation than the two-mass model.

B. Vocal tract load

The same experiment was performed with an acoustical load of an actual vocal tract of a female speaker pronouncing vowel /a/. F1 and F1′ of this vocal tract are calculated to be 874 and 1014 Hz, respectively. The result of the two-dimensional model is shown in Fig. 5 (a). The result of the two-mass models is shown in Fig. 5 (b).

We can observe similarities with the previous experiment. The two-dimensional model is capable of continuing self-excitation beyond F1. The sound frequency increases smoothly over F1 and falls into the capacitive region. On the other hand, the two-mass model again has a jump in the frequency when it reaches F1. In this case, however, the frequency jump is between 842 and 1106 Hz, and no self-excited oscillation can be generated in the entire capacitive region.

V. Discussions

The circular motion of the vocal fold is schematically illustrated in the left panel of Fig. 6. This section clarifies that this motion makes possible the self-sustained oscillation both with an inductive and capacitive vocal tract load.

The glottis is maximally opened at phase 2 and completely closed near phase 4. Therefore, a glottal area wave-

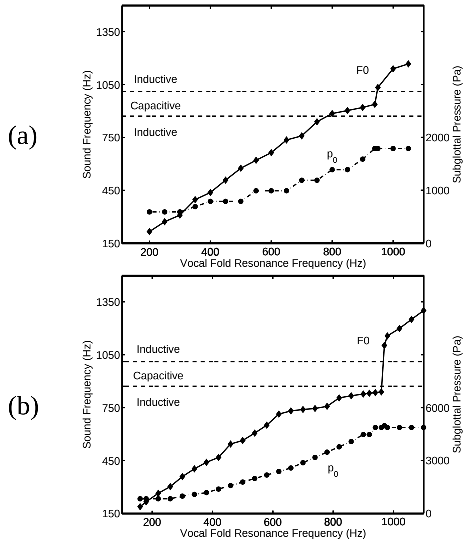


Figure 5: /a/ vowel simulation. (a) Two-dimensional model. (b) Two-mass model.

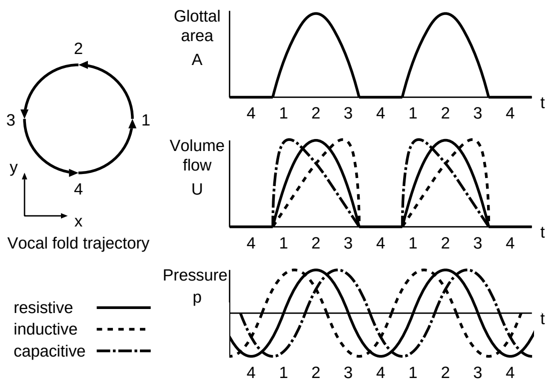


Figure 6: Schematic vocal fold trajectory and waveforms of glottal area, volume flow and pressure at the vocal tract entrance. Solid lines are for the case where the acoustic load is resistive (F0 ≈ F1), dashed lines are for the inductive load (F0 < F1) and dash-dot lines are for the capacitive load (F0 > F1).

form $A(t)$ shown in the upper-right panel is generated. When F0 is well below F1, the vocal tract acoustic load is inductive. In this case, according to [12], the volume flow $U(t)$ has a waveform with the slower rise and the abrupt fall as indicated by the dashed line in the middle-right panel. Because the phase of pressure $p(t)$ at the vocal tract entrance leads $U(t)$ up to 90 degree, $p(t)$ has a waveform as plotted by the dashed line in the lower-right panel. We then find that the glottal pressure, which is not very far from $p(t)$, pushes the vocal fold at phase 1 and sucks it at phase 3. Because the direction of the pressure is the same as that of the velocity of the vocal fold, the pressure becomes a driving force. This is the same mechanism for the one-mass vocal fold model to maintain self-sustained oscillation.

When F0 is close to F1, the vocal tract acoustic load becomes large and resistive. In this case, volume flow and pressure waveforms become symmetrical as indicated by the solid line. In this case, the glottal pressure does not drive the oscillation. Instead, $p(t)$ pushes the vocal fold upstream at phase 2 and downstream at phase 4 and becomes a driving force. This mechanism works even if the acoustic load becomes capacitive when F0 exceeds F1. The waveforms of volume flow and pressure are depicted in the dash-dot line. Pressure $p(t)$ takes its maximum between phase 2 and 3 and its minimum between phase 4 and 1. This causes a force in the direction of the vocal fold velocity.

Acknowledgment: This research was conducted as part of 'Research on Human Communication' with funding from the National Institute of Information and Communications Technology.

References

[1] J. Sundberg, *The science of the singing voice* (Northern Illinois Univ. Press., DeKalb, 1987) pp. 124–129.

[2] E. Joliveau, J. Smith and J. Wolfe, "Tuning of vocal tract resonance by sopranos," Nature **427** 116 (2004).

[3] E. Joliveau, J. Smith and J. Wolfe, "Vocal tract resonances in singing: The soprano voice," J. Acoust. Soc. Am. **116** 2434–2439 (2004).

[4] K. Ishizaka and J.L. Flanagan, "Synthesis of voiced sounds from a two-mass model of the vocal cords," Bell Syst. Tech. J. **51**(6), 1233–1268 (1972).

[5] S. Adachi and M. Sato, "Trumpet sound simulation using a two-dimensional lip vibration model," J. Acoust. Soc. Am. **99** 1200–1209 (1996).

[6] S. Adachi and M. Sato, "Time-domain simulation of sound production in the brass instrument," J. Acoust. Soc. Am. **97** 3850–3861 (1995).

[7] S. Adachi and J. Yu, "Two-dimensional model of vocal fold vibration for sound synthesis of voice and soprano singing," J. Acoust. Soc. Am. **117** 3213–3224 (2005).

[8] J. Liljencrants, "A translating and rotating mass model of the vocal folds," STL Quarterly Progress and Status Report, 1, Speech Transmission Laboratory (Royal Institute of Technology (KTH), Stockholm, Sweden), 1–18 (1991).

[9] J.L. Flanagan and K. Ishizaka, "Computer model to characterize the air volume displaced by the vibrating vocal cords," J. Acoust. Soc. Am. **63** 1559–1565 (1978).

[10] T. Baer, "Observation of vocal fold vibration: Measurement of excised larynges," in *Vocal folds physiology*, edited by K.N. Stevens and M. Hirano (University of Tokyo Press, Tokyo 1981), pp. 119–133.

[11] A.E. Rosenberg, "Effect of glottal pulse shape on the quality of natural vowels," J. Acoust. Soc. Am. **49**, 583–590 (1971).

[12] M. Rothenberg, "Acoustic interaction between the glottal source and the vocal tract," in *Vocal folds physiology*, edited by K.N. Stevens and M. Hirano (University of Tokyo Press, Tokyo 1981), pp. 305–323.