Effect of Unsteady Glottal Flow on the Speech Production Process

Hideyuki Nomura, Tetsuo Funada

Division of Electrical Engineering and Computer Science, Kanazawa University, Japan
nomu@t.kanazawa-u.ac.jp

Abstract

The purpose of the present study is to clarify the effects of unsteady glottal flow on the phonation. We numerically simulate the speech production process within the larynx and the vocal tract based on our proposed glottal sound source model. The simulation shows amplitude and waveform fluctuations in pressure within the larynx caused by unsteady fluid motion. In order to investigate the unsteady motion effects on the phonation, the coefficient of variation (CV) of amplitude and harmonic-to-noise ratio (HNR) in terms of measures of fluctuations are estimated. The CV and the HNR indicate the greatest fluctuation near the glottis, although the CV and the HNR do not show the fluctuation faraway from the glottis.

Index Terms: unsteady glottal flow, amplitude fluctuation, waveform fluctuation, coefficient of variation, harmonic-to-noise ratio

1. Introduction

A better understanding of complex phenomena within the larynx is of interest from biomechanics and speech processing standpoints. Recently, with improvements in computing power, a numerical simulation has become an efficient tool for clarifying the speech production process [1, 2, 3]. However, little information with respect to the effects of unsteady flow within the larynx on the speech production process is available.

In a previous study [4], we numerically simulated glottal flow based on a rigid wall model of the larynx and reported that, rather than being a steady symmetric laminar flow, the glottal flow is an unsteady asymmetric flow. In order to simulate the voiced sound, we have proposed a numerical glottal sound source model, which is constructed from a two-dimensional unsteady flow model and a distributed parameter model for the vocal cords [5, 6]. However, the effects of unsteady motion of the glottal flow on the voiced sound have not yet been investigated.

The purpose of the present study is to clarify the effects of the unsteady flow on pressure wave within the larynx and vocal tract. In order to investigate the unsteady motion effects on the phonation, the coefficient of variation of amplitude and harmonic-to-noise ratio in terms of measures of fluctuation are estimated using numerical simulations.

2. Model and computational method

A two-dimensional larynx model in the coronal (z-x) plane is shown in Fig. 1 (a). We assume that the configuration is uniform in the y direction, and that the initial configuration is symmetric about the glottal midline (z-axis). The vocal tract attached to the larynx is approximated by a uniform rigid duct.

The pressure difference between the inflow boundary \( \Gamma_1 \) and the outflow boundary \( \Gamma_2 \) generates flows within the larynx.

The interaction of the flow and vibrations of vocal folds generates the glottal sound source.

2.1. Mechanical vocal folds model

In order to consider vocal cord vibrations, we introduce a dynamical model as the boundary condition on the surfaces of the vocal cords \( \Gamma_3 \) and \( \Gamma_4 \). The present model of vibrating vocal cords is based on a distributed parameter model proposed by Ikeda et al.[7].

A vocal fold can be divided into two tissue layers with different mechanical properties: a cover layer and a body layer. The cover layer is assumed to be an elastic cover with the effective mass of vocal fold. In order to take the mechanical properties of vocal folds into account, the elastic cover is supported by small mechanical elements having nonlinear springs and dampers.

Figure 1 (b) shows the proposed vocal folds model. In Fig. 1 (b), there are only three mechanical elements on each side. In the present simulation, however, a greater number of elements, about 100 elements, are located on each surface of the larynx.

In the present study, in order to simplify the analysis, we restrict vocal fold vibrations in the lateral direction x. The vibration of vocal fold is governed by the equation of motion which consists of the fluid force, i.e., the pressure and the viscous stress of the flow, the restring force of the spring, the viscous drag force of the damper induced by the viscosity of the vocal fold tissue, and the shear force of the elastic cover [5, 6].

2.2. Glottal flow model

The fluid motion is assumed to be an unsteady two-dimensional compressible viscous fluid. The fluid analysis is based on boundary fitted coordinates along the surface of the larynx. Since the governing equations of flow are nonlinear partial differential equations, solving these equations analytically subject to initial and boundary conditions is extremely complicated. Instead, we employ a numerical computation method, MacCormack’s finite difference scheme [8] based on the arbitrary Lagrangian-Eulerian (ALE) method [9], for the speech production process.

2.3. Numerical simulation of speech production process

The speech production is simulated based on a moving boundary problem by alternately solving the flow equations and the motion equations of the vocal folds vibrations. A differential grid system moving with the wall boundary is employed to discretize the fluid domain. In order to determine an instantaneous larynx shape, vibrations of the mechanical elements of vocal fold are computed using the fourth-order Runge-Kutta method. Based on the instantaneous vocal fold displacement, we update
the boundary shape of the larynx and the differential grids in the fluid domain at every time step.

Speech production is simulated in the time domain as follows:

1. Initialize the larynx shape, differential grids, fluid motion, and vocal cord vibration.
2. Specify boundary conditions for fluid analysis.
3. Calculate the glottal flow.
4. Calculate the vibration of the vocal cords on the basis of the fluid force obtained by glottal flow analysis.
5. Update the boundary shape of the larynx and the differential grids.
6. Update time by one step and return to 2.

The details of the computational method are presented in previous studies [5, 6].

### 3. Results and discussion

#### 3.1. Conditions of simulation

Young’s modulus $E$, the viscosity $\nu$, the volume density $\rho_v$, and the effective thickness $h$ of the vocal fold tissue mainly determine the mechanical properties of the vocal fold vibration. Their typical values are listed in Table 1 according to the measured data [3, 10, 11, 12]. In the present study, the left and right vocal folds parameters are symmetric.

For the initial conditions, suppose that the air in the entire fluid space is uniform and at rest, and the vocal folds do not have strain and tension. The initial shape of the larynx is symmetric about the $z$ axis.

A pressure function $p_{L}(t)$ which is correspond to lung pressure is applied to the inflow boundary $\Gamma_1$ in Fig. 1, where $t$ is the time. The lung pressure function is defined as

$$p_{L}(t) = \begin{cases} P_{L0} & \text{for } 0 < t \leq t_r, \\ \frac{P_{L0}}{2} \left(1 - \cos \left(\frac{\pi t}{t_r}\right)\right) & \text{for } t > t_r. \end{cases} \quad (1)$$

Here, $P_{L0}$ and $t_r$ are the steady state value of the lung pressure and the rise time of the lung pressure, respectively. A non-reflecting characteristic boundary condition [13] is imposed at the outflow boundary $\Gamma_2$ in order to minimize acoustic reflection. The calculated instantaneous displacements and velocities of the vocal folds are applied to the boundaries (the surfaces of the larynx and the vocal tract) $\Gamma_3$ and $\Gamma_4$ at every moment.

#### 3.2. Results of simulation

Figure 2 shows examples of the left and right vocal folds vibrations, $d_{\text{left}}$ and $d_{\text{right}}$, and the pressure $p$ at different distance $z$ from the glottis. In the vocal folds vibrations, the solid and dotted lines of vocal fold vibration indicate the upper and lower vibrations of vocal folds, respectively.

The upper and lower lips of vocal folds vibrate with a phase difference. With increasing of lung pressure, amplitude and waveform fluctuations are observed. The fluctuation of pressure is largest at $z = 20$ mm. These fluctuations are caused by
3.3. Amplitude fluctuation

In order to investigate the amplitude fluctuations, we extract amplitude (peak-to-peak) sequences of pressure, shown in Fig. 3 (a), from the waveforms. From the extracted amplitude sequence in the range of \( t = 20 \) to 100 ms, the coefficient of variation (CV) of the amplitude sequence, which is the ratio of the standard deviation \( \sigma \) of the amplitude sequence to the average value \( m \) of the amplitude sequence, that is, \( CV = \sigma/m \), are obtained. Figure 4 shows the CV of pressure amplitude at different distance from the glottis. The solid, dotted, and dashed lines denote the fitting curves obtained using a linear function of the lung pressure.

The result indicates that the CVs of amplitude increase with lung pressure. The CV downstream of the glottis \( z = 20 \) mm is the greatest, and is of the order of \( 10^{10}\% \). On the other hand, the CV faraway from the glottis \( z = 160 \) mm is the smallest and is of the order of \( 10^{10}\% \).

3.4. Waveform fluctuation

In order to investigate the waveform fluctuation, the harmonic-to-noise ratio (HNR) is estimated. The HNR is a measure of the cycle to cycle similarity of the waveform, and is defined as the ratio between the energy of the periodic component to the energy of the aperiodic component in a wave [14]. An average wave is determined by taking the average of succession of period sequence, which is extracted based on a waveform of vocal folds vibration (glottal gap width), shown in Fig. 3 (b). The energy of the aperiodic component, i.e., noise, is the mean energy of the difference between the individual periods and the average waveform.

The original wave \( f(t) \) can be considered as the concatenation of the waves \( f_i(\tau) \) from each pitch period \( (i = 1, 2, \cdots, M \) and \( M \) is the number of samples). The average wave is estimated as

\[
 f_A(\tau) = \frac{1}{M} \sum_{i=1}^{M} f_i(\tau).
\]

For the calculating \( f_A(\tau) \), we assume that \( f_i(\tau) \) is equal to zero in the interval between \( T_i \) and \( T \), where \( T_i \) and \( T \) are the duration of the \( i \)-th period and the maximum period, respectively. The wave \( f_A(\tau) \) is defined over the interval from \( \tau = 0 \) to \( T \).

The energy of the periodic component is defined as

\[
 H = M \int_{0}^{T} f_A^2(\tau) d\tau.
\]

The noise wave in the \( i \)-th pitch period is equal to \( f_i(\tau) - f_A(\tau) \). The energy of the aperiodic component, i.e., noise, is
defining as

\[ N = \sum_{i=1}^{M} \int_{0}^{T_i} (f_i(\tau) - f_A(\tau))^2 \, d\tau. \]  

(4)

Then, the HNR is estimated as \( H/N \).

Figure 5 shows the HNR of the pressure wave at different distance from the glottis. The solid, dotted, and dashed lines denote the fitting curves obtained using a linear function of the lung pressure. The result indicates that the HNRs of the pressure decrease with lung pressure. The HNR of the pressure at \( z = 160 \) mm is roughly independent of the lung pressure and remains constant. The HNR of the pressure downstream of the glottis \( z = 20 \) mm is the smallest, and that faraway from the glottis \( z = 160 \) mm is the largest.

At the lung pressure of 800 Pa, which corresponds to the value for an ordinary conversation level, the HNRs are in the range of 10 to 18 dB. These values are agreement with the HNRs of 7 to 25 dB for real speech data [14].

4. Discussion and Conclusions

Unsteady motions of vortices generate a fluctuating wave within the larynx and vocal tract [4]. For pressure within the larynx at \( z = 20 \) mm, the CV is the largest, and the HNR is the smallest. On the other hand, faraway distance from the glottis \( z = 160 \) mm, the CV is the smallest, and the HNR is the largest. These results suggest that the unsteady motion of the glottal flow is restricted near the glottis and does not greatly affect speech waves radiated from the mouth for non-pathological speech organs with symmetric physical parameters.

5. Acknowledgements

The present research was supported in part by a Grant-in-Aid for Young Scientists (B) (No. 18700171) from the Japan Society for the Promotion of Science.

6. References


