A Simulation Study on the Effect of Glottal Boundary Conditions on Vocal Tract Formants

Yasufumi Uezu, Tokihiko Kaburagi
Kyushu University, 4-9-1 Shiobaru, Minami-ku, Fukuoka, Japan
uezu8223@gmail.com, kabu@design.kyushu-u.ac.jp

Abstract
In the source-filter theory of speech production [1], it is assumed that the source mechanism and the vocal-tract filter are independent. Also, the complete closure of the glottis is considered to be a boundary condition of the glottis. However, in an actual utterance, the glottis is opened and closed with the quasi-periodic self-excited vibration of the vocal folds by the expiratory flow. Therefore, boundary conditions of the glottis also change over time. Besides, via the utterance conditions, opening and closing pattern of the glottis may also be altered in various ways. This change of the boundary conditions of the glottis may also influence the coupling extent between the vocal tract and the lower part of the glottis. Thus, in an actual utterance, the assumption of glottal closure in the source-filter theory is not strictly satisfied. Acoustic features of the glottis and subglottis are considered to affect vocal tract formants. In this study, we investigated how differences in the glottal boundary conditions affect vocal tract formants by speech synthesis simulation using speech production model. We synthesized five Japanese vowels using the speech production model in consideration of the source-filter interaction. This model consisted of the glottal area polynomial model and the acoustic tube model in the concatenation of the vocal tract, glottis, and the subglottis. From the results, it was found that the first formant frequency was affected more strongly by the boundary conditions, and also found that the open quotient may give the formant stronger effect than the maximum glottal area. Furthermore, it was found that the open quotient may give the formant stronger effect than the maximum glottal area.

1. Introduction
In the source-filter theory of speech production [1], it is assumed that the source mechanism and the vocal-tract filter are independent. Also, the complete closure of the glottis is considered to be a boundary condition of the glottis. However, in an actual utterance, the glottis is opened and closed with the quasi-periodic self-excited vibration of the vocal folds by the expiratory flow. Therefore, boundary conditions of the glottis also change over time. Besides, via the utterance conditions, opening and closing pattern of the glottis may also be altered in various ways. This change of the boundary conditions of the glottis may also influence the coupling extent between the vocal tract and the lower part of the glottis. Thus, in an actual utterance, the assumption of glottal closure in the source-filter theory is not strictly satisfied. Acoustic features of the glottis and subglottis are considered to affect the vocal tract formants.

Barney et al. [2] conducted a model experiment that combined the glottal model which opened and closed periodically and the uniform rectangular tube made of acrylic. If the glottal opening area were set to be time-invariant, it was found that the first formant frequency was affected more strongly by the boundary conditions, and also found that the open quotient may give the formant stronger effect than the maximum glottal area.

2. Physical model of speech production
2.1. Polynomial model of the glottal area
Titze and Story [3] investigated the effect of the cross-sectional area and the length of the lower vocal tract part (e.g. epilarynx, piriform) on vocal tract formants. On the other hand, the aim of our study is to investigate how the maximum glottal area and the glottal open quotient influence vocal tract formants. For the purpose, it is required that the glottal model behaves constantly for given glottal parameters, without suffering the effect such as the acoustical feedback from vocal tract. Therefore, we introduced the polynomial model of the glottal area, instead of the vocal-fold physical model such as the n-mass model [4, 5]. Thus, it was possible to examine the influence of glottal parameters on the formants.

The time waveforms of the glottal area and the glottal volume flow had a proportional relationship. Using the polynomial model of the glottal volume flow by Rosenberg [6], time waveforms of the glottal volume flow and the glottal area \( G_A \) were represented as Eq. (1) and Eq. (2), respectively.

\[
G_A = \alpha (t^2 - t^4)G_M \quad (0 \leq t \leq 1) \tag{1}
\]

\[
U_g = G_A \sqrt{\frac{2P_0}{\rho}} \tag{2}
\]

Here, \( G_M \) was a maximum glottal area, \( \alpha \) was a coefficient for normalization, \( P_0 \) was lung pressure, and \( \rho \) was the air density. It should be noted that the time axis of \( G_A \) is relative. In fact, \( G_A \) was used for the speech synthesis by scaling its time axis.

In the speech synthesis simulation, the maximum glottal area and the open quotient (OQ) were given as vocalization parameters. By providing the OQ, it was possible to determine the percentage of an opening period to the fundamental period of the glottal opening and closing.
2.2. Acoustic tube model in the concatenation of the vocal tract, the glottis, and the subglottis

We introduced the acoustic tube model of Sondhi and Schroeter [7] as a physical model of the vocal tract. Through the acoustic tube model, the vocal tract was described as a multi-tube approximation by connecting a plurality of cylindrical tubes that had different cross-sectional areas. The relationship between input and output for the sound pressure and the volume velocity was expressed as the Eq. (3). Its propagation matrix and propagation coefficients were represented by Eq. (4).

\[
\begin{pmatrix}
A & B \\
C & D
\end{pmatrix} = \prod_{i=1}^{N} \begin{pmatrix}
A_i & B_i \\
C_i & D_i
\end{pmatrix}
\]

(3)

\[
A_i = \cosh(\sigma L_{gi}/c) \\
B_i = -(pc/S_{gi})\gamma(\sinh(\sigma L_{gi}/c)) \\
C_i = -(S_{gi}/pc)(\sinh(\sigma L_{gi}/c))/\gamma \\
D_i = \cosh(\sigma L_{gi}/c)
\]

(4)

\[
Z_{in} = \frac{P_{in}}{U_{in}} \quad H = \frac{U_{out}}{U_{in}}
\]

(5)

Here, \(L_{gi}\) is the length of the cylindrical tube, \(S_{gi}\) is the cross-sectional area of the cylindrical tube, \(\rho\) is sound velocity, \(\sigma\) and \(\gamma\) are frequency-dependent coefficient. \(i\) featured index numbers representing sections of the vocal tract.

Vocal tract input impedance \(Z_{in}\) and the vocal tract transfer function \(H\) were obtained from Eq. (5).

In this study, we used the consolidated acoustic tube model of the vocal tract, the glottis, and the subglottis based on the acoustic tube model described above. Figure 1 shows a schematic view of the connected acoustic tube model [8], \(A_0, B_0, C_0, D_0\), \(A_1, B_1, C_1, D_1\), and \(A_2, B_2, C_2, D_2\) represented the propagation coefficients of the subglottis, the vocal tract, and the glottis, respectively. \(Z_0\), \(Z_g\), and \(Z_1\) represented respectively the input impedance of the subglottis, the glottis, and the vocal tract. \(Z_t\) is the radiation impedance of the lips [10], and \(Z_p\) is the termination impedance of the subglottis [11]. Input impedances \(Z_1\), \(Z_0\), and \(Z_g\), along with the vocal tract transfer function \(H_1\), were calculated using Eq. (6), (7), (8), and (9).

\[
Z_0 = -\frac{A_0 Z_p + B_0}{C_0 Z_p + D_0}
\]

(6)

\[
Z_1 = \frac{D_1 Z_0 - B_1}{A_1 - C_1 Z_0}
\]

(7)

\[
Z_g = -\frac{B_2 - A_2 Z_0}{D_2 - C_2 Z_0}
\]

(8)

\[
H_1 = \frac{1}{A_1 - C_1 Z_0}
\]

(9)

Finally, the transfer function of the connected acoustic tube model, \(H_{ct}\), was able to be calculated according to Eq. (10).

\[
H_{ct} = \frac{H_1 Z_g}{Z_1 + Z_g}
\]

(10)

Figure 1: A schematic view of the consolidated acoustic tube model of the vocal tract, the glottis, and the subglottis based on the acoustic tube model.

2.3. Speech synthesis simulation procedure

Figure 2 shows a flow chart of the simulation procedure of speech synthesis in this study. First, the glottal area waveform was calculated using vocalization parameters: the fundamental frequency, OQ, and the maximum glottal area. Further, the subglottal input impedance \(Z_{in}\), the vocal tract input impedance \(Z_t\), and the vocal tract transfer function \(H_1\) were calculated using area functions of the subglottis and the vocal tract. Next, the following calculation was repeated while updating the time. The glottal area was calculated from the \(G_A\) at a certain time \(t\), and then, the glottal input impedance \(Z_g\) and the glottal volume flow \(U_g\) were computed. Next, using \(Z_g\), \(Z_0\), \(Z_1\), and \(H_1\), the entire route transfer function \(H_{ct}\) was calculated. After that, the volume flow of the lips was computed by convolving \(U_g\) and the impulse response of \(H_{ct}\). Finally, the speech signal was obtained by applying the differential filter as the radiation impedance of the lip to the volume flow gained from repeat calculation.

3. Experiment

We synthesized speech signals under different glottal boundary conditions using the speech production model described in section 2. The vocal tract cross-sectional area function data of five Japanese vowels /a/, /i/, /u/, /e/, and /o/ were used for speech synthesis. This data was extracted from the three-dimensional vocal tract MRI imaging data according to Story et al. [9]. MRI imaging data was obtained from a Japanese adult male as a subject. The cross-sectional area data of the bronchi and lungs in Weibel [12] were used as the subglottal area function. The length of the acoustic tube representing the glottis was set to 3 mm. The fundamental frequency was set to 100 Hz, and
the glottal open quotient was set from 0.4 to 0.8 at 0.2 intervals. The maximum glottal width was set from 0.25 mm to 3.00 mm at 0.25 mm intervals. The maximum glottal area was calculated by using the maximum glottal width. Therefore, the maximum glottal area used was from 4.25 mm to 51 mm. By combining these vocalization parameters, 60 types of speech signals per one vowel were synthesized. Other synthesis parameters were set as follows: the lung pressure $P_L = 8 \text{ cmH}_2\text{O}$, the air density $\rho = 1.184 \times 10^{-3} \text{ g/cm}^3$, the sound velocity $c = 34630 \text{ cm/sec}$, the time length of the synthesized sound = 0.2 sec, the sampling frequency = 48 kHz, and the DFT score = $2^{14}$.

From all synthesized speech, the vocal tract resonance characteristics were analyzed by applying linear prediction analysis. Then, formant frequencies were analyzed by applying the peak picking. The LPC order was set to 12. In this analysis, a regular 100 msec interval of the synthetic speech was cut out. Then, downsampling to 10 kHz and the pre-emphasis were applied. Finally, window processing was done using the Hamming window.

4. Results and discussion

4.1. The first formant frequency

Figure 3 and figure 4 show the results of the first and the second formant frequencies of the synthesized vowel /a/. Figure 5 through figure 8 show the results of the first formant frequen-

cies of synthesized vowels /i/, /u/, /e/, and /o/. The solid red line in each figure shows the “vocal-tract” first formant frequency $F_1$ ($F_2'$ in figure 4). Red dashed and dotted lines in each figure show the range of $F_1 \pm 10\%$, $F_1 \pm 20\%$ respectively ($F_2'$ in figure 4).

In all vowels, the first formant frequency tended to be higher as compared to the first formant frequency $F_1'$. In particular, it was found that $F_1'$ increased more than 20% for the vowel /i/ and more than 10% for the vowel /o/. As for the maximum glottal width and formant frequency, it was found that the first formant frequency tended to be more increased when the maximum glottal width became larger in the cases of vowels /i/, /u/, and /o/. Regarding the OQ and the formant frequency, it was found that the first formant frequency tended to be more increased when the OQ became higher. It was considered that the glottal boundary condition became a more open state as the maximum glottis width and the OQ increased. Thus, the vocal tract resonance approached open tube resonance. As a result, the first formant frequency was increased.

4.2. The subglottal impedance and the vocal tract transfer function

There is also a possibility that the peak frequency of the subglottal impedance affects the first and the second formant frequencies. Figure 9 shows the subglottal impedance used in this study. Note that the subglottal impedance had lower peaks at
The results of the second formant frequency were different from those of the first formant; a consistent trend was not observed. However, it was suggested that the effect of an open quotient on the second formant depended on the value of the maximum glottal width. It was found that the second formant frequency for vowels /a/ and /o/ were strongly influenced by the maximum glottal width. The second formants of these vowels were about 1500 Hz in the vowel /a/ and about 1200 Hz. The first formant frequency in vowels was found to be relatively strongly influenced by the maximum glottal width and the OQ. The first formants of these vowels from about 400 Hz to about 500 Hz existed in the vicinity of the first peak of the subglottal impedance. Therefore, such first formant frequencies changed significantly because of the effect of the subglottal impedance with changes in glottis boundary conditions.

The results of the second formant frequency were different from those of the first formant; a consistent trend was not observed. However, it was suggested that the effect of an open quotient on the second formant depended on the value of the maximum glottal width. It was found that the second formant frequency for vowels /a/ and /o/ were strongly influenced by the maximum glottal width and the OQ. The second formants of these vowels were about 1500 Hz in the vowel /a/ and about 1100 Hz in the vowel /o/, which existed in the vicinity of the first peak of the subglottal impedance.

The impact of the subglottal impedance peaks on vocal tract formants was confirmed to vary depending on glottis boundary conditions. Various values of the glottal open quotient and the maximum glottal width were given as glottal parameters for synthesis. The results, the influence of the glottal boundary conditions on formants was confirmed to vary depending on vocal tract area function. The first formant frequency in the consolidated acoustic tube model was confirmed to be higher than that of the vocal tract in all vowels. Regarding glottis boundary conditions, it was found that the maximum glottal width tends to have the more dominant influence on formant frequencies than the open quotient. It was suggested that the changes in formant frequencies occurred because the vocal tract resonance approached the open tube resonance when the glottal area was wider, or the open quotient was higher. It was also suggested that the change of formant frequencies occurred because the peak of the subglottis impedance was influenced more strongly when the glottal area was wider, or the open quotient was higher.

### 5. Summary

In this paper, we studied the effect of glottal boundary conditions on vocal tract formants using computer simulations of speech production. We synthesized five Japanese vowels by using a speech production model composed of a polynomial model of the glottal area and an acoustic tube model that is a concatenation of the vocal tract, the glottis, and the subglottis. Various values of the glottal open quotient and the maximum glottal width were given as glottal parameters for synthesis. From the results, the influence of the glottal boundary conditions on formants was confirmed to vary depending on vocal tract area function. The first formant frequency in the consolidated acoustic tube model was confirmed to be higher than that of the vocal tract in all vowels. Regarding glottis boundary conditions, it was found that the maximum glottal width tends to have the more dominant influence on formant frequencies than the open quotient. It was suggested that the changes in formant frequencies occurred because the vocal tract resonance approached the open tube resonance when the glottal area was wider, or the open quotient was higher. It was also suggested that the change of formant frequencies occurred because the peak of the subglottis impedance was influenced more strongly when the glottal area was wider, or the open quotient was higher.
6. References


