A MODEL TO REPRESENT PROPAGATION AND RADIATION OF HIGHER-ORDER MODES FOR 3-D VOCAL-TRACT CONFIGURATION

Kunitoshi MOTOKI and Hiroki MATSUZAKI

Faculty of Engineering, Hokkai-Gakuen University
S-26,W-11, Sapporo, JAPAN
E-mail: motoki@eli.hokkai-s-u.ac.jp

ABSTRACT

For the representation of acoustic characteristics of three-dimensional vocal-tract shapes, it is necessary to consider the effects of higher-order modes. This paper proposes an acoustic model of the vocal-tract which incorporates the coupling of the higher-order modes, including both propagative and evanescent modes. A cascaded structure of acoustic tubes connected asymmetrically is introduced as a physical approximation of the vocal-tract. The acoustic characteristics, which are dependent not only on the vocal-tract area function but also on the vocal-tract configuration, can be investigated by the proposed model. Preliminary results of numerical computations for relatively simple configurations suggest that additional resonances at frequencies above 4.3kHz are formed by the propagative higher-order modes, while those at frequencies below 3 kHz are influenced by the evanescent higher-order modes. These results are also confirmed by the FEM simulations.

1. INTRODUCTION

Visualizations of the acoustic field in vocal-tracts obtained by acoustic measurements[1] and FEM simulations [2] indicate that the assumption of plane wave propagation in the vocal-tracts is not sufficient for the representation of acoustic characteristics of three-dimensional vocal-tract shapes even at relatively lower frequencies around 3 kHz.

In this paper, a parametric acoustic model of the speech production system with the propagation and radiation of higher-order modes is proposed. A coupling of higher-order modes, including evanescent modes, is expressed using a mode decomposition technique. A cascaded structure of acoustic tubes connected asymmetrically is introduced as an approximation of the geometry of the vocal-tract. In order to facilitate the mode decomposition of the acoustic field in the model, each tube is assumed to have a rectangular cross-sectional shape with an aspect ratio appropriate for the corresponding part of the vocal-tract.

The proposed model consists of three parts; sound source, vocal-tract, and radiation parts, which can be considered as an extension of the well-known one-dimensional model since the proposed model is exactly reduced to the one-dimensional model when no higher-order modes are taken into account. For each of these parts, electrical equivalent circuits are presented. One higher-order mode can be regarded as one transmission line with a characteristic impedance and a propagation constant corresponding to those of that higher-order mode. Thus, multi-wire lines connected with ideal transformers are introduced as the equivalent circuits. The radiation boundary is also represented by ideal transformers terminated by radiation impedances that can be defined for each higher-order mode.

For the evaluation of acoustic characteristics of the proposed model, the transfer function between a source volume velocity and the sound-pressure at a distant position is derived from the viewpoint of the effective radiation power. Since the transfer function is expressed in terms of parameters of higher-order modes, influences of each higher-order mode can be examined frequency by frequency.

The preliminary results obtained by using relatively simple tube geometries show that many resonances due to the propagative higher-order modes appear in the higher frequencies. Zeros also appear in the higher frequencies. Moreover, an interesting fact is that resonance frequencies below the cut-off frequency of the first higher-order mode are lowered by the existence of the evanescent higher-order modes. This result implies that the asymmetrical geometry of the connected tubes can be regarded as a certain extension of length for waves to travel. These results are also confirmed by Finite Element Method simulations.

For real vocal-tracts, such asymmetrical geometry may be formed especially in the region anterior to the tongue tip. The acoustic characteristics, which are dependent not only on the vocal-tract area function but also on the vocal-tract configuration, can be investigated by the proposed model.

2. 3-D PARAMETRIC MODEL

2.1 Higher-order modes in a tube

The 3-D acoustic field in a uniform rectangular tube can be represented in the infinite series of higher-order modes. A sound-pressure $p(x, y, z)$, z being the direction of the tube axis, and z direction particle velocity $v_z(x, y, z)$ are expressed as follows [3].

$$
\begin{aligned}
p(x,y,z) = & \sum_{N=0}^{\infty}(a_N e^{-\gamma_N z} + b_N e^{\gamma_N z})\phi_N(x,y) \\
\approx & \ \phi^T\{\mathbf{D}(-z)\mathbf{a} + \mathbf{D}(z)\mathbf{b}\} \\
v_z(x,y,z) \approx & \ \phi^T\mathbf{Z}_C^{-1}\{\mathbf{D}(-z)\mathbf{a} - \mathbf{D}(z)\mathbf{b}\}
\end{aligned}
$$

where N stands for the number of the higher-order mode, $\gamma_N, \phi_N(x,y)$ are the propagation constant and normal function. In the above matrix notation, the infinite series are truncated to a certain value N_{max}, and $\mathbf{a}, \mathbf{b}$, ϕ , $\mathbf{Z}_C$ and $\mathbf{D}(z)$ are defined as,

$$\mathbf{a} = [a_0, a_1, \cdots, a_{N_{max}}]^T \quad , \quad \mathbf{b} = [b_0, b_1, \cdots, b_{N_{max}}]^T$$
$$\phi = [\phi_0(x,y), \phi_1(x,y), \cdots, \phi_{N_{max}}(x,y)]^T$$
$$\mathbf{Z}_C = jk\rho c(\mathrm{diag}[\gamma_0, \gamma_1, \cdots, \gamma_{N_{max}}])^{-1}$$
$$\mathbf{D}(z) = \mathrm{diag}[e^{\gamma_0 z}, e^{\gamma_1 z}, \cdots, e^{\gamma_{N_{max}} z}]$$

where k, ρ and c are wave number, air density and sound speed, respectively. For the representation of propagation, scattering and radiation of higher-order modes, root power waves are used, which are defined as,

$$\mathbf{w}_a = (2\mathbf{Z}_C)^{-1/2}\mathbf{a} \quad , \quad \mathbf{w}_b = (2\mathbf{Z}_C)^{-1/2}\mathbf{b}.$$

2.2 Vocal-tract

When two tubes are connected asymmetrically, the coupling of the higher-order modes can be expressed in terms of a scattering matrix $\mathbf{S}_w$ which represents the relationship between the incident waves $(\mathbf{w}_{a_1}, \mathbf{w}_{b_2})$ and the reflected waves $(\mathbf{w}_{b_1}, \mathbf{w}_{a_2})$, and is the function of the geometry of the junction [4],

$$\left[\begin{array}{c} \mathbf{w}_{b_1} \\ \hline \mathbf{w}_{a_2} \end{array}\right] = \overbrace{\left[\begin{array}{c|c} \mathbf{S}_w^{11} & \mathbf{S}_w^{12} \\ \hline \mathbf{S}_w^{21} & \mathbf{S}_w^{22} \end{array}\right]}^{\mathbf{S}_w} \left[\begin{array}{c} \mathbf{w}_{a_1} \\ \hline \mathbf{w}_{b_2} \end{array}\right] \begin{array}{l} \}\mathrm{N}_1 \\ \}\mathrm{N}_2. \end{array}$$

The suffix denotes the tube number. N_1 and N_2 are the total number of modes that should be considered in tube 1 and 2. For the better representation of the acoustic field near the junction, the ratio of N_1 and N_2 should be chosen in accord with that of the sectional size of the two tubes [4, 5, 6]. In that case, $\mathbf{S}_w^{21}, \mathbf{S}_w^{12}$ are not square matrices. Let us assume that $\mathbf{S}_w$ is properly obtained by using the proper values of N_1 and N_2. Then, it can be justifiably assumed that the modes of very high order are localized in the vicinity of the junction, and have almost no influence on the transmission of the acoustic power in the tube with a length of L_z. Thus, if we extract the upper left parts of $\mathbf{S}_w^{11}, \mathbf{S}_w^{21}, \mathbf{S}_w^{12}$ and $\mathbf{S}_w^{22}$, with a size of $N_T \times N_T$, where $N_T < \min(N_1, N_2)$, the incident and reflected waves can be expressed in terms of a transfer scattering matrix $\mathbf{T}_w$ which is a suitable form to represent a cascaded structure of several tubes.

$$\left[\begin{array}{c} \mathbf{w}_{a_1} \\ \mathbf{w}_{b_1} \end{array}\right] = \mathbf{T}_w \left[\begin{array}{c} \mathbf{w}_{a_2} \\ \mathbf{w}_{b_2} \end{array}\right]$$

The electrical equivalent circuits and signal flow of the junction are shown in figure 1. Multi transmission lines corresponding to each higher-order mode are connected by an ideal transformer. Let the length of the tube 1 be L_{z_1}, and rewrite the forward and backward wave components on the end of the sound source side of the tube as $\mathbf{w}_{a_1}$ and $\mathbf{w}_{b_1}$. Then the forward and backward wave components on the end of the sound source side of each tube can be expressed in the following transfer scattering matrix $\mathbf{T}_w^{(1)}$.

$$\left[\begin{array}{c} \mathbf{w}_{a_1} \\ \mathbf{w}_{b_1} \end{array}\right] = \overbrace{\left[\begin{array}{cc} \mathbf{D}(L_{z_1}) & \mathbf{0} \\ \mathbf{0} & \mathbf{D}(-L_{z_1}) \end{array}\right] \mathbf{T}_w}^{\mathbf{T}_w^{(1)}} \left[\begin{array}{c} \mathbf{w}_{a_2} \\ \mathbf{w}_{b_2} \end{array}\right]$$

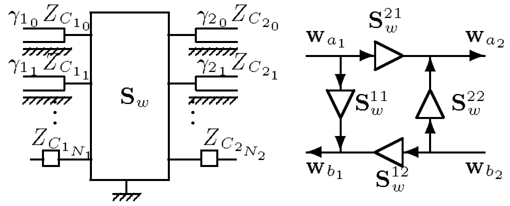

Figure 1: Equivalent circuit and signal flow for junction.

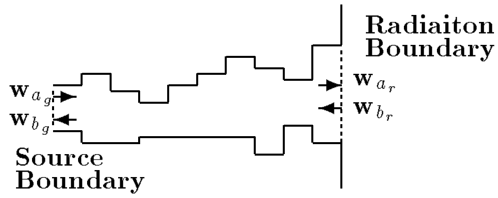

Figure 2: Model configuration.

Thus, if the vocal-tract is approximated as a cascaded structure of N_s sections of rectangular tubes as shown in figure 2, the forward and backward wave components at the sound source region ($\mathbf{w}_{a_g}$ and $\mathbf{w}_{b_g}$) and those at the radiation region ($\mathbf{w}_{a_r}$ and $\mathbf{w}_{b_r}$) can be represented as follows.

$$\left[\begin{array}{c} \mathbf{w}_{a_g} \\ \mathbf{w}_{b_g} \end{array}\right] = \prod_{i=1}^{N_s} \mathbf{T}_w^{(i)} \left[\begin{array}{c} \mathbf{w}_{a_r} \\ \mathbf{w}_{b_r} \end{array}\right]$$

2.3 Sound source

Assume that the particle velocity distribution $v_g(x,y)$ is given on the surface Ω_g of the sound source region of the first tube. The mode decomposition representation of $v_g(x,y)$ is

$$v_g(x,y) \approx \phi_1^T \mathbf{Z}_{C_1}^{-1}\{\mathbf{a}_1 - \mathbf{b}_1\}.$$

If we define a vector $\mathbf{v}_g$ as,

$$\mathbf{v}_g = \int\int_{\Omega_g} \phi_1 v_g(x,y) dS$$

then we get,

$$\mathbf{w}_{a_g} = (\mathbf{Z}_{C_1}/2)^{1/2}\mathbf{v}_g + \mathbf{w}_{b_g}$$

where,

$$\mathbf{w}_{a_g} = (2\mathbf{Z}_{C_1})^{-1/2}\mathbf{a}_1 \ , \ \mathbf{w}_{b_g} = (2\mathbf{Z}_{C_1})^{-1/2}\mathbf{b}_1.$$

An electrical equivalent circuit and signal flow for a source with higher-order modes are shown in figure 3.

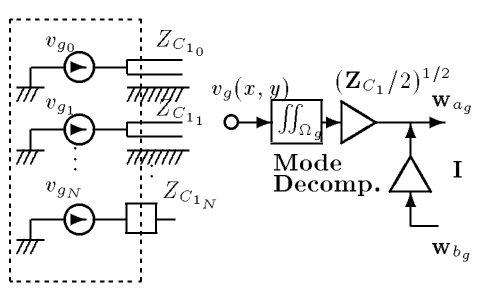

Figure 3: Equivalent circuit and signal flow for source.

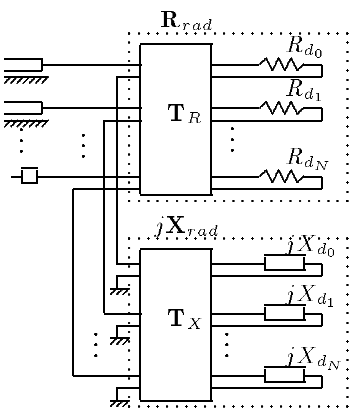

Figure 4: Equivalent circuit and signal flow for radiation.

2.4 Radiation

Radiation is a very important factor for determining the resonance characteristics of the vocal-tract. Based on acoustic measurements, we have already reported that the radiation powers originating from the propagative higher-order modes can be greater than those of plane waves [7]. Here, we assume that the last section of the vocal-tract is open on an infinite plane baffle. The scattering matrix that gives the relationship between the forward and backward wave components at the radiation region is given by,

$$\mathbf{S}_w^{rr} = \mathbf{Z}_{C_{N_s}}^{1/2}(\mathbf{Z}_{rad} + \mathbf{Z}_{C_{N_s}})^{-1}(\mathbf{Z}_{rad} - \mathbf{Z}_{C_{N_s}})\mathbf{Z}_{C_{N_s}}^{-1/2}$$

where, $\mathbf{Z}_{rad}$ is a radiation impedance matrix derived by Muehleisen [4] and is defined as,

$$\mathbf{Z}_{rad} = \frac{jk\rho c}{2\pi}\iiiint \phi\ (x,y)\phi\ ^T(x',y')\frac{e^{-jkr}}{r}dx'dy'dxdy,$$
$$r = \sqrt{(x-x')^2 + (y-y')^2}$$

$\phi\ (x,y)$ is a vector of normal functions of the last section, and the integrand is integrated over the open end region. $\mathbf{Z}_{rad}$ represents modal radiation impedance. The first diagonal element of $\mathbf{Z}_{rad}$ is the ordinary radiation impedance (per unit area) that is widely used in traditional speech production models. Since $\mathbf{Z}_{rad}$ is a symmetric matrix, $\mathbf{Z}_{rad}$ can be decomposed as,

$$\begin{aligned}\mathbf{Z}_{rad} &= \mathbf{R}_{rad} + j\mathbf{X}_{rad}\\ &= \mathbf{T}_R^T\mathbf{R}_d\mathbf{T}_R + j\mathbf{T}_X^T\mathbf{X}_d\mathbf{T}_X\\ &\quad (\mathbf{T}_R^T\mathbf{T}_R = \mathbf{I}\ \ ,\ \ \mathbf{T}_X^T\mathbf{T}_X = \mathbf{I})\end{aligned}$$

where $\mathbf{R}_d$ and $\mathbf{X}_d$ are diagonal matrices composed of eigen values of $\mathbf{R}_{rad}$ and $\mathbf{X}_{rad}$, and $\mathbf{T}_R$ and $\mathbf{T}_X$ are matrices composed of eigen vectors of $\mathbf{R}_{rad}$ and $\mathbf{X}_{rad}$. Thus, from classical circuit theory, an electrical equivalent circuit for radiation with higher-order modes can be realized as shown in figure 4.

2.5 Transfer function

Define a vocal-tract transfer function as,

$$H = K|p_{far}/u_g|$$

where u_g and p_{far} are a volume velocity of a sound source and a sound pressure at a distant position from the open end, respectively. K is a frequency-independent constant that is needed for H to be dimensionless. The effective radiation power $W_{p_{rad}}$ can be obtained by integrating active acoustic intensities over the open end.

$$\begin{aligned}W_{p_{rad}} &= \int\int \mathrm{Re}\{\bar{p}(x,y)v_z(x,y)/2\}dxdy\\ &= \mathrm{Re}\{\mathbf{w}_{a_r}^*(\mathbf{I}+\mathbf{S}_w^{rr})^*(\mathbf{Z}_{C_{N_s}}^*)^{1/2}\times\\ &\quad \mathbf{Z}_{C_{N_s}}^{-1/2}(\mathbf{I}-\mathbf{S}_w^{rr})\mathbf{w}_{a_r}\}\end{aligned}$$

At a distant position, sound waves are assumed to be plane waves. Then $|p_{far}|^2$ should be proportional to $W_{p_{rad}}$. Thus, using another frequency-independent constant K_w, we can write ,

$$|p_{far}| = K_w\sqrt{W_{p_{rad}}}$$

u_g is easily obtained by integrating $v_g(x,y)$ over Ω_g. Thus we can compute H from the above 2 equations.

3. PRELIMINARY RESULTS

The transfer functions for relatively simple configurations, as shown in figure 5, are computed from the proposed model. The driving source region is located at the center of the first section with a size of 0.5cm width × 1.48cm height. The distribution of $v_g(x,y)$ is uniform over this source region.

Tube size:
$L_{x_1} = L_{x_3} = 4.0$ cm, $L_{x_2} = 5.66$ cm
$L_{y_1} = L_{y_3} = 1.48$ cm, $L_{y_2} = 2.09$ cm
$L_{z_1} = L_{z_2} = L_{z_3} = 5.5$ cm

Condition:

A Symmetry: Figure 5(a)	Plane waves only
B Symmetry: Figure 5(a)	With higher-order modes
C Asymmetry: Figure 5(b)	With higher-order modes

Figure 6 shows the transfer characteristics of each condition. The cut-off frequencies of the first and last section are indicated by vertical lines at 4.33 and 8.66kHz, and those of the second section are indicated by arrows. It is seen that some additional resonances in the higher frequencies above 4.3kHz are formed by the propagative higher-order modes. There are three resonances in the lower frequencies below 3 kHz. These frequencies are listed in Table 1. The values in parentheses are the rate of change relative to condition **A**. It is clearly seen that the resonance frequencies are lowered by the existence of the evanescent higher-order modes. This result implies that the asymmetrical geometry of the connected tubes can be regarded as a certain extension of length for waves to travel.

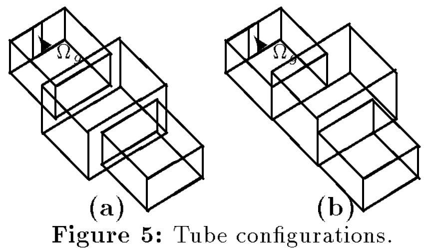

Figure 5: Tube configurations.

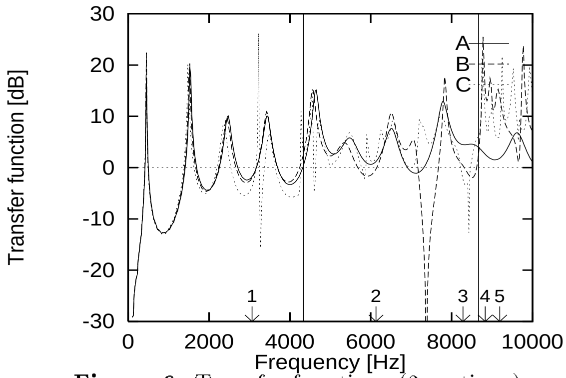

Figure 6: Transfer functions (3 sections).

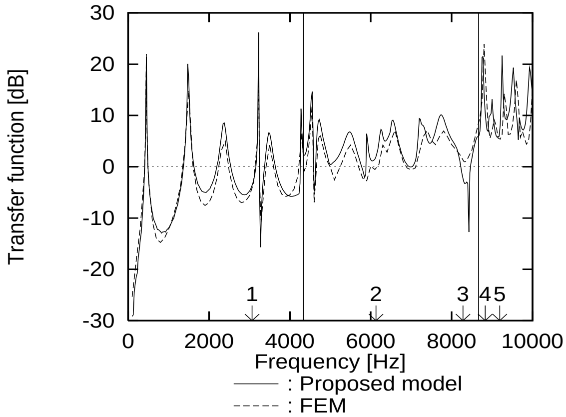

Figure 7: Transfer functions (comparison with FEM).

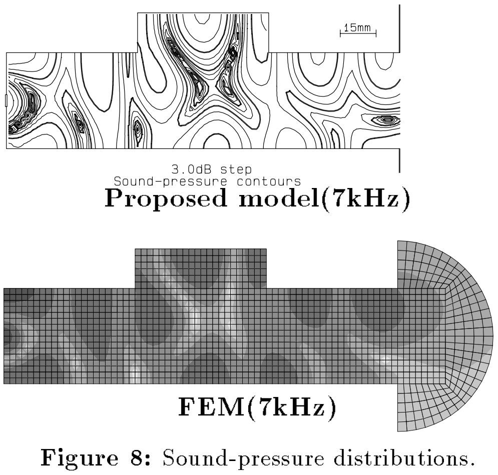

Figure 8: Sound-pressure distributions.

Table 1: Peak frequencies [Hz].

	f_1	f_2	f_3
A	455	1536	2467
B	452 (-0.7)	1516 (-1.3)	2443 (-1.0)
C	449 (-1.3)	1484 (-3.4)	2366 (-4.1)

Figure 7 shows the simulation results of figure 5(b) by FEM with an additional hemispherical radiation space attached to the open end [2]. The characteristics obtained by the proposed model are also drawn as a solid line. And figure 8 presents the sound pressure distributions at 7 kHz obtained by the proposed method and FEM simulation, and shows very good agreement. At other frequencies, the sound pressure distributions obtained by the two different methods show good agreement as well. From these figures, we can confirm that the proposed model validly represents the acoustic characteristics of three-dimensional vocal-tract configuration.

4. CONCLUSION

The proposed model is an extension of the traditional one-dimensional model of speech production, and is valid to represent the steady-state frequency characteristics of three-dimensional configuration. A direct simulation in the time domain, however, is not available since delay-free loops due to the evanescent modes appear in the model. For the improvement of the proposed model, mode decomposition of bent and branch tubes should be investigated as a future project.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Nobuhiro Miki, Hokkaido University, for his continuous support and suggestions for our research.

Part of this work was supported by a Grant-in-Aid for Scientific Research (09750522) from the Ministry of Education, Japan.

REFERENCES

[1] K.Motoki, N.Miki and N.Nagai, “Measurement of sound-pressure distribution in replicas of the oral cavity,” J. Acoust. Soc. Am., **92**,5,2577-2585(1992).

[2] H.Matsuzaki, N.Miki and Y.Ogawa, “FEM analysis of sound wave propagation in the vocal tract with 3-D radiational model,” J. Acoust. Soc. Jpn. (E), **17**, 163–166(1996).

[3] P.M.Morse and K.U.Ingard,*Theoretical acoustics* (McGraw-Hill, New York, 1968).

[4] R.T.Muehleisen, “Reflection, radiation, and coupling of higher order modes at discontinuities in finite length rigid walled rectangular ducts,” Ph.D thesis, Pennsylvania State Univ. (1996).

[5] R.Mittra and S.W.Lee, *Analytical techniques in the theory of guided waves* (Macmillan, New York, 1971), chap.2.

[6] H.Hudde and U.Letens, “Scattering matrix of a discontinuity with a nonrigid wall in a lossless circular duct,” J. Acoust. Soc. Am., **78**, 1826-1837 (1985).

[7] K.Motoki and N.Miki, “Effects of higher-order modes on sound radiation characteristics at mouth opening,” Proc. ASA and ASJ third joint meeting, 2pSC42,895-898(1996).