A MODAL PARAMETRIC METHOD FOR COMPUTING ACOUSTIC CHARACTERISTICS OF THREE-DIMENSIONAL VOCAL TRACT MODELS

Kunitoshi MOTOKI*, Pierre BADIN**, Xavier PELORSON**, and Hiroki MATSUZAKI*

*Department of Electronics and Information Engineering, Hokkai-Gakuen University
S-26,W-11, Sapporo, 064-0926, JAPAN
E-mail: {motoki,matsu}@eli.hokkai-s-u.ac.jp

**Institut de la Communication Parlée, UMR CNRS 5009, INPG-Université Stendhal
46, Avenue Félix Viallet, 38031 Grenoble, FRANCE
E-mail: {badin,pelorson}@icp.inpg.fr

ABSTRACT

A model for calculating the acoustic characteristics of three-dimensional vocal tract configuration is presented. A cascaded structure of rectangular acoustic tubes, connected asymmetrically with respect to their axes, is introduced as an approximation of the real vocal tract geometry. Both propagative and evanescent higher-order modes are considered in these tubes. The number of higher-order modes can be selected independently in each tube, which significantly decreases the instability of computation caused by the evanescent higher-order modes. Calculation results for two configurations are discussed.

1 INTRODUCTION

Recent development of techniques for the observation of speech organs such as MRI allows us to obtain accurate descriptions of vocal tract shape. Although 3D shapes of the vocal tract are available, 1D theories are still used to compute acoustic characteristics of the vocal tract. A reason maybe lies in the complexity of computation of the resonance characteristics of 3D shapes.

Numerical computation techniques such as the Finite Element Method (FEM) [1] or the Transmission Line Matrix (TLM) method [2] have been applied to compute the acoustic characteristics of 3D vocal tract models. The results obtained emphasize the large influence of the vocal tract shape details upon the transfer function. These methods, however, require a large amount of computation, and are not suitable for the purpose of speech synthesis.

This paper presents a parametric method to compute the acoustic characteristics of the 3D vocal tract model, in order to achieve the reduction of the computation, and to explore the vocal tract acoustics that can not be represented by the traditional 1D model. A cascaded structure of acoustic tubes, connected asymmetrically with respect to their axes, is introduced as an approximation of the vocal tract geometry. Each tube is assumed to have a rectangular cross-sectional shape whose geometry (size and axis position) can be determined from MRI data. The 3D acoustic field in each tube is represented in terms of higher-order modes. A mode-matching technique is then used to establish a mode coupling at the junctions between tubes. In the proposed method, both propagative and evanescent higher-order modes are considered in each tube, since each section is often not long enough for the evanescent modes to decay away.

Considering several evanescent higher-order modes sometimes causes computational difficulties related to the numerical precision. As a matter of fact, in a previous report [3], the number of higher-order modes taken into account in each tube was necessarily constant; this was a major drawback when constrictions were present since the evanescent higher-order modes in the narrow tube often caused computational instability. In the proposed method, the number of higher-order modes can be selected independently in each tube. In particular, only plane waves may be considered for narrow tubes and several higher-order modes can be taken into account for wider tubes. The flexibility in the selection of the number of the higher-order modes in each tube increases the computation stability significantly, while also reducing computational time.

Calculation results for two configurations, a 5-section configuration approximating an occlusion at the teeth, and 36-section configuration based on MRI data, are discussed.

2 3D VOCAL TRACT MODEL

2.1 Mode expansion and coupling

Vocal tracts are approximated by cascaded structures of rectangular acoustic tubes, as shown in figure 1. The 3D acoustic field in each tube can be represented in infinite series of higher-order modes. The sound-pressure $p(x,y,z)$, z being the direction of the tube axis, and the z-direction particle velocity $v_z(x,y,z)$ in each tube are expressed as:

$$\begin{aligned} p(x,y,z) &= \sum_{m,n=0}^{\infty} (a_{mn}e^{-\gamma_{mn}z} + b_{mn}e^{\gamma_{mn}z})\phi_{mn}(x,y) \\ &\approx \boldsymbol{\phi}^T(x,y)\{\mathbf{D}(-z)\mathbf{a} + \mathbf{D}(z)\mathbf{b}\} \\ v_z(x,y,z) &\approx \boldsymbol{\phi}^T(x,y)\mathbf{Z}_C^{-1}\{\mathbf{D}(-z)\mathbf{a} - \mathbf{D}(z)\mathbf{b}\} \end{aligned} \quad (1)$$

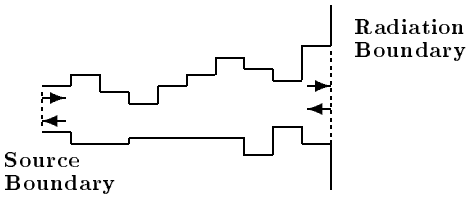

Figure 1. Example of vocal tract model.

where m and n stand for the numbers of the higher-order modes in x and y directions, γ_{mn} and $\phi_{mn}(x,y)$ are the propagation constant and normal function (eigen function), respectively. γ_{mn}'s are imaginary numbers for propagative modes and real numbers for evanescent modes. Neither loss factors due to the viscosity and the heat conductivity of air, nor wall vibration effects are included. In the matrix notation in eq. (1), the infinite series are truncated to a certain value. $\mathbf{a}, \mathbf{b}$ and $\boldsymbol{\phi}\ (x,y)$ are column vectors composed of a_{mn}, b_{mn} and $\phi_{mn}(x,y)$, respectively. $\phi_{mn}(x,y)$'s are chosen to have the following orthogonal property:

$$\frac{1}{S}\int_S \boldsymbol{\phi}\ (x,y)\boldsymbol{\phi}\ (x,y)^T dS = \mathbf{I} \tag{2}$$

where S is the area of the tube, and $\mathbf{I}$ a unit matrix. $\mathbf{a}$ and $\mathbf{b}$ are determined by the boundary conditions at the both ends of each tube. $\mathbf{D}(z)$ and $\mathbf{Z}_C$ are defined as:

$$\mathbf{D}(z) = \mathrm{diag}[\exp(\gamma_{mn} z)]\ ,\ \mathbf{Z}_C = jk\rho c(\mathrm{diag}[\gamma_{mn}])^{-1} \tag{3}$$

where k, ρ and c are wave number, air density and sound speed, respectively. A modal sound-pressure vector $\mathbf{P}$ and a modal particle velocity vector $\mathbf{V}$ can be defined as:

$$\begin{aligned} \mathbf{P} &= \mathbf{D}(-z)\mathbf{a} + \mathbf{D}(z)\mathbf{b} \\ \mathbf{V} &= \mathbf{Z}_C^{-1}\{\mathbf{D}(-z)\mathbf{a} - \mathbf{D}(z)\mathbf{b}\} \end{aligned} \tag{4}$$

If each component of $\mathbf{P}$ and $\mathbf{V}$ is regarded as a voltage and a current at position z, each higher-order mode can be represented by an equivalent electrical transmission line. Hereafter, a subscript i is used to represent the variables in the section i. By using a mode matching technique, the mode coupling at the junction between the sections i and $i+1$ can be expressed as follows[4]:

$$\begin{aligned} \mathbf{P}_i &= \boldsymbol{\Phi}_{i,i+1}\mathbf{P}_{i+1} \\ \boldsymbol{\Phi}_{i,i+1}^T\mathbf{V}_i &= \mathbf{V}_{i+1} \end{aligned} \tag{5}$$

where the coupling matrix $\boldsymbol{\Phi}_{i,i+1}$ is calculated as:

$$\boldsymbol{\Phi}_{i,i+1} = \frac{1}{S_i}\int_{S_i} \boldsymbol{\phi}_{\ i}(x,y)\boldsymbol{\phi}^{\ T}_{i+1}(x,y)dS \tag{6}$$

S_i is assumed to be smaller than that of the section $i+1$. If the smaller area is for section $i+1$, all suffixes i and $i+1$ should just be exchanged. Equation (5) indicates that the coupling coefficients matrix $\boldsymbol{\Phi}_{i,i+1}$ can be simply regarded as a matrix representing the transformation ratio of a multi-port ideal transformer in an equivalent electrical circuit.

2.2 Impedance transformation

Acoustic transfer functions and sound-pressure distributions in the vocal tract can be calculated from input impedance matrices at each section. Starting with a given load impedance matrix, which is often a radiation impedance matrix, an impedance transformation is necessary to obtain the input impedance matrices for all sections. Kergomard [4] presented a "propagating modes" method for the impedance transformation where all evanescent modes are terminated at the junction. In this section, the formulation of the impedance transformation suitable for the 3D vocal-tract model is presented. It can handle short sections with asymmetrical connection between the adjacent tubes. It should be noted that the evanescent higher-order modes do not propagate. However, they can influence the resonance characteristics through the mode coupling between plane waves and the evanescent modes at the junction. Moreover, if two junctions are located very closely, the evanescent modes can be related to a power transmission. Thus we consider the evanescent modes either as "terminated" at the junction or as "related to the power transmission" in the tube as illustrated in figure 2.

2.2.1 Radiation impedance matrix

A radiation impedance matrix is the first impedance for starting the impedance transformation. The generalized modal radiation impedance matrix $\mathbf{Z}_{rad}$ for a rectangular opening with the higher-order modes is given by [5] as:

$$\begin{aligned} \mathbf{Z}_{rad} &= [Z_{mn,pq}] \\ Z_{mn,pq} &= \frac{jk\rho c}{2\pi S}\int_S\int_S \phi_{mn}(x,y)\phi_{pq}(x',y')\frac{e^{-jkr}}{r}dS'dS, \\ r &= \sqrt{(x-x')^2+(y-y')^2} \end{aligned} \tag{7}$$

where S is the area of the last section corresponding to the mouth. Note that $Z_{mn,pq}$ represents the modal radiation impedance between modes (m,n) and (p,q). $Z_{00,00}$ is the radiation impedance used in the plane wave theory.

2.2.2 Basic transformation

Tube Input impedances looking toward loads at the right and left ends of the section i are defined as:

$$\mathbf{P}_i^{(\mathrm{R})} = \mathbf{Z}_i^{(\mathrm{R})}\mathbf{V}_i^{(\mathrm{R})}\ \ ,\ \ \mathbf{P}_i^{(\mathrm{L})} = \mathbf{Z}_i^{(\mathrm{L})}\mathbf{V}_i^{(\mathrm{L})} \tag{8}$$

where superscripts $^{(\mathrm{R})}$ and $^{(\mathrm{L})}$ are used to denote the quantities at the right (lip side) and left (glottis side) ends of the tube. From eq. (4), the impedance transformation from $\mathbf{Z}_i^{(\mathrm{R})}$ to $\mathbf{Z}_i^{(\mathrm{L})}$ is easily obtained as:

$$\mathbf{Z}_i^{(\mathrm{L})} = (\mathbf{D}_{C_i}\mathbf{Z}_i^{(\mathrm{R})} + \mathbf{D}_{S_i}\mathbf{Z}_{C_i})(\mathbf{D}_{S_i}\mathbf{Z}_i^{(\mathrm{R})} + \mathbf{D}_{C_i}\mathbf{Z}_{C_i})^{-1}\mathbf{Z}_{C_i} \tag{9}$$

where $\mathbf{D}_{C_i} = \{\mathbf{D}_i(L_i) + \mathbf{D}_i(-L_i)\}/2$ and $\mathbf{D}_{S_i} = \{\mathbf{D}_i(L_i) - \mathbf{D}_i(-L_i)\}/2$, L_i being the length of the section i.

Junction The relation between impedance matrices $\mathbf{Z}_i^{(\mathrm{R})}$ and $\mathbf{Z}_{i+1}^{(\mathrm{L})}$ is given from the well-known circuit theory as:

$$\mathbf{Z}_i^{(\mathrm{R})} = \boldsymbol{\Phi}_{i,i+1}\mathbf{Z}_{i+1}^{(\mathrm{L})}\boldsymbol{\Phi}_{i,i+1}^T \tag{10}$$

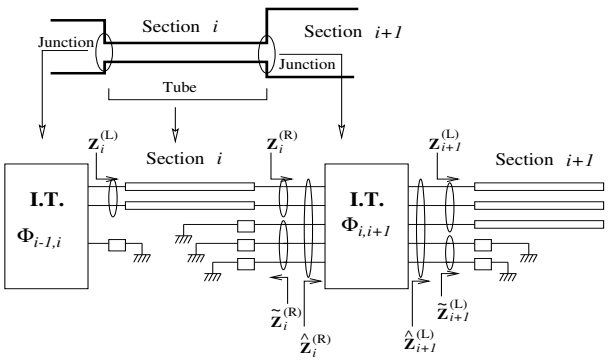

Figure 2. Equivalent electrical circuit.

2.2.3 Transformation with terminations

We assume that the coupling coefficient matrix $\mathbf{\Phi}_{i,i+1}$ is square. The number of modes selected for the calculation of $\mathbf{\Phi}_{i,i+1}$ can be large. It may be, however, limited to 5 or 6 as described in the next section. Some of these modes are considered for transmission, while the others are terminated with their characteristic impedances as illustrated in figure 2. Some input impedance matrices are also defined in figure 2. The problem to solve for the impedance transformation is to express $\mathbf{Z}_i^{(\mathrm{L})}$ in terms of $\tilde{\mathbf{Z}}_i^{(\mathrm{R})}$,$\mathbf{Z}_{i+1}^{(\mathrm{L})}$ and $\hat{\mathbf{Z}}_{i+1}^{(\mathrm{L})}$. $\tilde{\mathbf{Z}}_i^{(\mathrm{R})}$ and $\tilde{\mathbf{Z}}_{i+1}^{(\mathrm{L})}$ are simply diagonal matrices composed of characteristic impedances used for termination of the ideal transformer. $\hat{\mathbf{Z}}_{i+1}^{(\mathrm{L})}$ is written as,

$$\hat{\mathbf{Z}}_{i+1}^{(\mathrm{L})} = \begin{bmatrix} \mathbf{Z}_{i+1}^{(\mathrm{L})} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{Z}}_{i+1}^{(\mathrm{L})} \end{bmatrix} \tag{11}$$

Then from eq. (10), we get,

$$\hat{\mathbf{Z}}_i^{(\mathrm{R})} = \mathbf{\Phi}_{i,i+1} \hat{\mathbf{Z}}_{i+1}^{(\mathrm{L})} \mathbf{\Phi}_{i,i+1}^T \tag{12}$$

$\hat{\mathbf{Z}}_i^{(\mathrm{R})}$ can be decomposed into sub matrices as:

$$\hat{\mathbf{Z}}_i^{(\mathrm{R})} = \begin{bmatrix} \hat{\mathbf{Z}}_{i,11}^{(\mathrm{R})} & \hat{\mathbf{Z}}_{i,12}^{(\mathrm{R})} \\ \hat{\mathbf{Z}}_{i,21}^{(\mathrm{R})} & \hat{\mathbf{Z}}_{i,22}^{(\mathrm{R})} \end{bmatrix} \tag{13}$$

where the size of $\hat{\mathbf{Z}}_{i,11}^{(\mathrm{R})}$ corresponds to the number of modes considered for transmission in the i-th section. Then $\mathbf{Z}_i^{(\mathrm{R})}$ is calculated as:

$$\mathbf{Z}_i^{(\mathrm{R})} = \hat{\mathbf{Z}}_{i,11}^{(\mathrm{R})} - \hat{\mathbf{Z}}_{i,12}^{(\mathrm{R})} \{\hat{\mathbf{Z}}_{i,22}^{(\mathrm{R})} + \tilde{\mathbf{Z}}_i^{(\mathrm{R})}\}^{-1} \hat{\mathbf{Z}}_{i,21}^{(\mathrm{R})} \tag{14}$$

Finally, $\mathbf{Z}_i^{(\mathrm{L})}$ is obtained by substituting $\mathbf{Z}_i^{(\mathrm{R})}$ into eq. (9). Repeating the above procedure section by section, the given radiation impedance matrix $\mathbf{Z}_{rad}$ in eq. (7) is transformed into the input impedance matrix of the first section.

2.3 Transfer function

Once all input impedance matrices are obtained, wave component vectors $\mathbf{a}$ and $\mathbf{b}$ at each section can be calculated using eqs. (4) and (5) recursively. Then sound-pressure distribution is obtained from eq. (1). It is also possible to compute the sound-pressure at an arbitrary far point by using the Rayleigh integral with the calculated vibrating pattern at the radiation end. However, a more convenient way to evaluate the resonance characteristics is to use the power that is actually radiated in the free space. Since the present model assumes no losses inside the vocal tract, the actual radiation power is equal to the total active intensities W_t on any arbitrary cross-section in tubes, which can be evaluated at the glottis section as:

$$W_t = \int_{S_1} \frac{1}{2} \mathrm{Re}\{p_1^* v_{z_1}\} dS = \frac{S_1}{2} \mathrm{Re}\{(\mathbf{Z}_1^{(\mathrm{L})} \mathbf{V}_1^{(\mathrm{L})})^* \mathbf{V}_1^{(\mathrm{L})}\} \tag{15}$$

where * denotes a complex conjugate, and $\mathbf{V}_1$ is the modal particle velocity vector of the given pattern of the source vibration. Then the sound-pressure at a far point can be considered to be proportional to $\sqrt{W_t}$. Thus the transfer function H of the proposed model is evaluated by,

$$H \propto \frac{\sqrt{W_t}}{U_G} \tag{16}$$

where U_G is a source volume velocity.

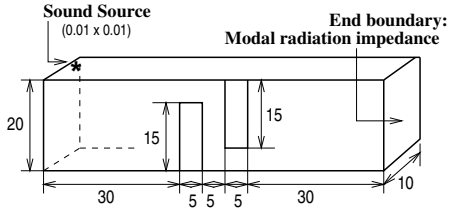

Figure 3. 5-section configuration.

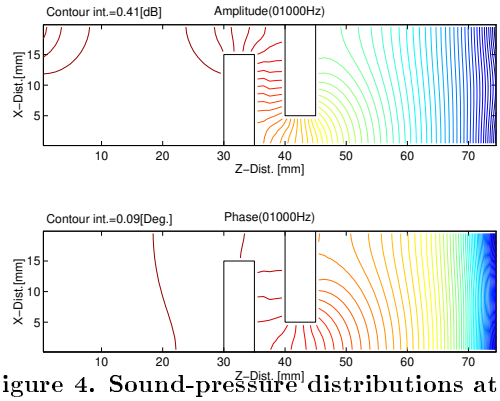

Figure 4. Sound-pressure distributions at 1kHz.

3 RESULTS

3.1 Configuration with 5-section

A 5-section configuration is used to imitate the occlusion at the teeth as illustrated in figure 3. A tiny square sound source (0.01 mm $\times$ 0.01mm) is located at 0.5 mm apart from the upper right corner. In the narrow sections 2 and 4 only plane waves are considered for the transmission and the first 2 higher-order modes (1,0) and (2,0) are considered for the mode coupling at each junction. Computed sound-pressure distribution at 1 kHz is shown in figure 4, both amplitude and phase contours being presented. Note that these higher-order modes are all evanescent at 1 kHz. The superposition of the plane waves and of these two higher-order modes makes the waves travel almost in the vertical direction in the occlusion section. As a result, the acoustic field becomes just like *another plane waves* being propagated in the vertical direction. This result indicates that this method can be also used to determine the parameters of the 1D vocal tract model, such as estimating an equivalent length L_e and area S_e of the third section for the plane waves to propagate in the vertical direction. One simple way to estimate these values is to use the computed wave parameters for plane waves adjacent to the third section. The components for plane waves at the right end of the second section and the left end of the fourth section should be related by L_e and S_e with the plane wave theory as:

$$\begin{bmatrix} P_{2_{00}}^{(\mathrm{R})} \\ S_2 V_{2_{00}}^{(\mathrm{R})} \end{bmatrix} = \begin{bmatrix} \cos(kL_e) & j\frac{\rho c}{S_e}\sin(kL_e) \\ j\frac{S_e}{\rho c}\sin(kL_e) & \cos(kL_e) \end{bmatrix} \begin{bmatrix} P_{4_{00}}^{(\mathrm{L})} \\ S_4 V_{4_{00}}^{(\mathrm{L})} \end{bmatrix} \tag{17}$$

where the suffix "$_{00}$" is used to represent plane wave components of modal pressure and particle velocity vectors. L_e and S_e are easily obtained from this equation. The resultant L_e is almost constant up to 6 kHz while S_e is gradually decreasing with the ascent of the frequency. The average values up to 5 kHz are $L_e = 1.74$cm and $S_e = 0.56$cm^2 while the axis length and cross-sectional area of the third section are 0.5 cm and 2 cm^2. As easily imagined from figure 4, the occlusion has the effect of extending the path for

equivalent plane wave propagation at low frequencies. This lengthening can be evaluated quantitatively with the proposed method. This result is coherent with measurements [6] performed using circular tubes with similar "occlusion-like" shapes. Even though the proposed method is to represent the 3D acoustic field in the asymmetrical tube configuration, it can be also used to establish a proper area function for "occlusion-like" shape for the traditional 1D model.

3.2 Configuration based on MRI data

In order to study the differences between the plane wave and the higher-order models, the 3D shape of a /∫/ measured by MRI has been converted into a 36-section configuration as illustrated in figure 5. Each tube is aligned to a common horizontal plane. The entire configuration is symmetrical with respect to the lateral direction, but asymmetrical in the vertical direction. A sound source is assumed to be located at the beginning of the first section. 5 higher-order modes (3 in lateral direction and 2 in vertical direction) are considered at each junction. Figure 6 shows the acoustic transfer characteristics defined in eq. (16) together with those obtained from the FEM simulation and from the 1D model. The peak frequencies obtained by the proposed method agree well with those from the FEM although a slight difference around a "zero" at 6.4 kHz is seen. The frequency of the zero tends to be more sensitive to the number of the higher-order modes than the peak frequencies. For peak frequencies up to 5 kHz, only a few higher-order modes are necessary to give the same peak frequencies as those of FEM. Comparison with the result of the 1D model suggests that the resonance frequencies are always lowered due to the evanescent higher-order modes. The rates of shift relative to the 1D model for the first 4 resonance frequencies are -3.8, -4.0, -3.1, and -4.3 %, respectively. Above the 5-th resonance frequency, the plane wave and the higher-order models have quite different transfer characteristics. The "zero" can appear around the first cut-off frequency in the lateral direction of the widest section. This may be understood using the analogy with side branches, since the higher-order modes in each section correspond to transmission lines in parallel as an equivalent electrical circuit.

4 CONCLUSION

It has been shown that the proposed model is valid to represent the steady-state frequency characteristics for 3D configurations of rectangular tubes used as a geometrical approximation of 3D vocal tracts. In particular, the proposed method presents the following advantages: (1) more accurate acoustic characteristics, compared to those obtained from the 1D modeling; (2) shorter computational time compared to FEM and/or TLM. These are useful features for the improvement of speech synthesis systems based on vocal tract models. It seems that the number of the modes needed to represent the 3D acoustic field for the rectangular geometry is not large. However, a criterion for the selection of the number of modes for the given configuration is not well established. Another possible problem is that unwanted artifacts might occur due to the use of the rectangular geometry. Further studies are needed with regard to the manner of geometrical approximation of real vocal tracts.

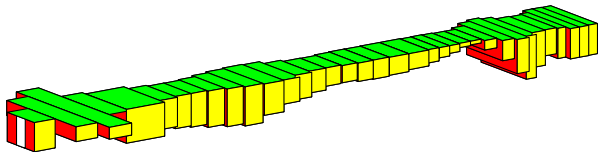

Figure 5. 36-section configuration (/∫/).

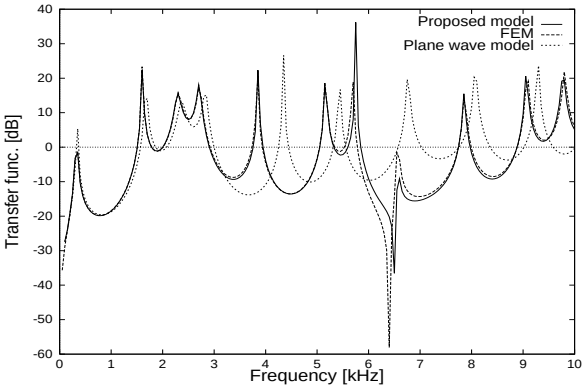

Figure 6. Transfer functions of 36-section configuration.

ACKNOWLEDGMENTS

Part of this work has been performed at ICP during the first author's stay as a visiting researcher from March 1999 to February 2000. The first author wish to thank to Dr. R.Laboissière, ICP, for discussing the acoustics with higher-order modes, and to Dr. N.Miki, Hokkaido Univ., for his valuable suggestions for this work. Part of this work has been supported by a research project of High-Tech Research Center, Hokkai-Gakuen Univ., and by CREST of JST (Japan Science and Technology).

REFERENCES

[1] Matsuzaki,H., Motoki,K. and Miki,N., "Effects of shapes of radiational aperture on radiation characteristics", Proc. ICSLP98, Sydney, Australia,Tu5D6,547-550(1998).

[2] El-Masri,S., Pelorson,X., Saguet,P., and Badin,P., "Development of the transmission line matrix method in acoustics applications to higher modes in the vocal tract and other complex ducts" Intl. J. Numerical Modelling, 11, 133-151(1998).

[3] Motoki,K. and Matsuzaki,H., "A model to represent propagation and radiation of higher-order modes for 3-D vocal-tract configuration", Proc. ICSLP98, Sydney, Australia,Fr1R14,3123-3126(1998).

[4] Kergomard,J.,"Calculation of discontinuities in waveguides using mode-matching method: an alternative to the scattering matrix approach",J. Acoustique, **4**, 111-138(1991).

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

[6] Motoki,K., Miki,N., and Nagai,N., "On the influence of discontinuous shapes near the lips", Proc. Acoust. Soc. Jpn. Autumn meeting, 165-166(1987).