COMPUTATION OF 3-D VOCAL TRACT ACOUSTICS BASED ON MODE-MATCHING TECHNIQUE

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

* Department of Electronics and Information Engineering, Hokkai-Gakuen University
S-26,W-11, Sapporo, 064-0926, JAPAN {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 {pelorson,badin}@icp.inpg.fr

ABSTRACT

A model for calculating the acoustic characteristics of 3-D vocal tract configuration is presented. A mode-matching technique and an impedance transformation are used to calculate the higher-order modes in this model where both propagative and evanescent higher-order modes are considered. A cascaded structure of rectangular acoustic tubes is introduced as an approximation of the real vocal tract geometry. The number of higher-order modes in each tube can be selected independently, which significantly decreases the instability of computation caused by the evanescent higher-order modes. Preliminary calculation results are discussed from the view point of the influences of the offset connection of tubes.

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. Numerical computation techniques such as the Finite Element Method (FEM) [1] or the Transmission Line Matrix (TLM) method [2] have been applied to investigate the detailed acoustic characteristics of 3-D 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 method to compute the acoustic characteristics of the 3-D 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 1-D model. A cascaded structure of rectangular acoustic tubes is introduced as an approximation of the vocal tract geometry. The geometry (size and axis position) of each tube can be determined from MRI data. The 3-D acoustic field in each tube is represented by the superposition of plane waves and several higher-order modes. A mode-matching technique and an impedance transformation are used to calculate the higher-order modes in this model where both propagative and evanescent higher-order modes are considered. In many applications relating to duct acoustics, the evanescent modes are not taken into account since these modes are localized near the junctions between tubes. However, the evanescent modes are important to represent the 3-D acoustic characteristics of the vocal tract model since each tube is not long enough for the evanescent modes to decay away.

Considering several evanescent modes sometimes causes computational difficulties related to the numerical precision. 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 computational stability significantly, while also reducing computational time.

Computational results are discussed from the view point of the influences of the offset connection. A 5-section configuration imitating an occlusion at the teeth is used to evaluate the path for plane wave to propagate in the occlusive area. The influences of the small perturbation of the position of tube axes are also shown using 36-section configurations from MRI data.

2. 3-D VOCAL TRACT MODEL

2.1. Mode expansion and coupling

A cascaded structure of rectangular acoustic tubes, connected asymmetrically with respect to their axes, is introduced as an approximation of the vocal tract geometry. The 3-D acoustic field in each tube can be represented in infinite series of higher-order modes. The sound-pressure $p(x,y,z)$, x being the direction of the tube axis, and the z-direction particle velocity $v_z(x,y,z)$ in each tube are expressed as:

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

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. 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. $\mathbf{a}$ 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:

$$\mathbf{D}(z) = \mathrm{diag}[\exp(\gamma_{mn} z)] \ , \ \mathbf{Z}_C = jk\rho c(\mathrm{diag}[\gamma_{mn}])^{-1} \quad (2)$$

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:

$$\begin{aligned} \mathbf{P} &= \mathbf{D}(-z)\mathbf{a} + \mathbf{D}(z)\mathbf{b} \\ \mathbf{V} &= \mathbf{Z}_C^{-1}\{\mathbf{D}(-z)\mathbf{a} - \mathbf{D}(z)\mathbf{b}\} \end{aligned} \quad (3)$$

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]:

$$\begin{aligned} \mathbf{P}_i &= \mathbf{\Phi}_{i,i+1}\mathbf{P}_{i+1} \\ \mathbf{\Phi}_{i,i+1}^T\mathbf{V}_i &= \mathbf{V}_{i+1} \end{aligned} \quad (4)$$

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

$$\mathbf{\Phi}_{i,i+1} = \frac{1}{S_i}\int_{S_i} \boldsymbol{\phi}_i(x,y)\boldsymbol{\phi}_{i+1}^T(x,y)dS \quad (5)$$

S_i is an area of the section i. Equation (4) indicates that the coupling coefficients matrix $\mathbf{\Phi}_{i,i+1}$ 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

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 3-D vocal-tract model is presented. 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. Thus we consider the evanescent modes either as *terminated* (lumped element) at the junction or as *connected*(line element) in the tube as illustrated in figure 1.

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:

$$\begin{aligned} \mathbf{Z}_{rad} &= [Z_{mn,pq}] \\ Z_{mn,pq} &= \frac{jk\rho c}{2\pi S}\int_S\int_S \phi_{mn}(x,y)\phi_{pq}(x',y')\frac{e^{-jkr}}{r}dS'dS, \\ r &= \sqrt{(x-x')^2+(y-y')^2} \end{aligned} \quad (6)$$

where S is the area of the last section corresponding to the mouth. $Z_{00,00}$ is the radiation impedance used in the plane wave theory.

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

$$\mathbf{P}_i^{(R)} = \mathbf{Z}_i^{(R)}\mathbf{V}_i^{(R)} \ , \ \mathbf{P}_i^{(L)} = \mathbf{Z}_i^{(L)}\mathbf{V}_i^{(L)} \quad (7)$$

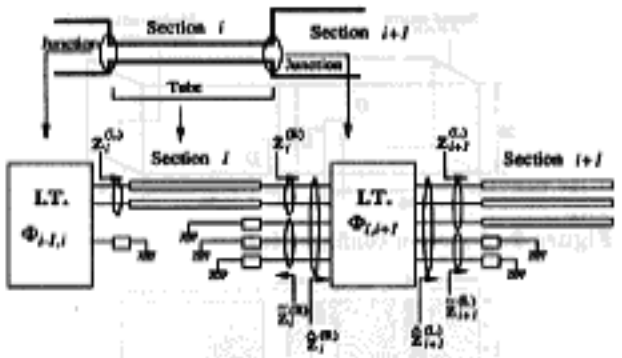


Figure 1: Equivalent electrical circuit.

where superscripts $^{(R)}$ and $^{(L)}$ are used to denote the quantities at the right (lip side) and left (glottis side) ends of the tube. From eq. (3), the impedance transformation by a tube (from $\mathbf{Z}_i^{(R)}$ to $\mathbf{Z}_i^{(L)}$) is easily obtained as:

$$\mathbf{Z}_i^{(L)} = (\mathbf{D}_{C_i}\mathbf{Z}_i^{(R)} + \mathbf{D}_{S_i}\mathbf{Z}_{C_i})(\mathbf{D}_{S_i}\mathbf{Z}_i^{(R)} + \mathbf{D}_{C_i}\mathbf{Z}_{C_i})^{-1}\mathbf{Z}_{C_i} \quad (8)$$

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. The impedance transformation by a junction (from $\mathbf{Z}_{i+1}^{(L)}$ to $\mathbf{Z}_i^{(R)}$) is given from the well-known circuit theory as:

$$\mathbf{Z}_i^{(R)} = \mathbf{\Phi}_{i,i+1}\mathbf{Z}_{i+1}^{(L)}\mathbf{\Phi}_{i,i+1}^T \quad (9)$$

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 1. Some input impedance matrices are also defined in figure 1. The problem to solve for the impedance transformation at each section is to express $\mathbf{Z}_i^{(L)}$ in terms of $\hat{\mathbf{Z}}_i^{(R)}, \mathbf{Z}_{i+1}^{(L)}$ and $\hat{\mathbf{Z}}_{i+1}^{(L)}$. $\hat{\mathbf{Z}}_i^{(R)}$ and $\hat{\mathbf{Z}}_{i+1}^{(L)}$ are simply diagonal matrices composed of characteristic impedances used for termination of the ideal transformer. $\hat{\mathbf{Z}}_{i+1}^{(L)}$ is written as,

$$\hat{\mathbf{Z}}_{i+1}^{(L)} = \begin{bmatrix} \mathbf{Z}_{i+1}^{(L)} & \mathbf{0} \\ \mathbf{0} & \hat{\mathbf{Z}}_{i+1}^{(L)} \end{bmatrix} \quad (10)$$

Then from eq. (9), we get,

$$\hat{\mathbf{Z}}_i^{(R)} = \mathbf{\Phi}_{i,i+1}\hat{\mathbf{Z}}_{i+1}^{(L)}\mathbf{\Phi}_{i,i+1}^T \quad (11)$$

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

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

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

$$\mathbf{Z}_i^{(R)} = \hat{\mathbf{Z}}_{i,21}^{(R)} - \hat{\mathbf{Z}}_{i,12}^{(R)}\{\hat{\mathbf{Z}}_{i,22}^{(R)} + \hat{\mathbf{Z}}_i^{(R)}\}^{-1}\hat{\mathbf{Z}}_{i,21}^{(R)} \quad (13)$$

Finally, $\mathbf{Z}_i^{(L)}$ is obtained by substituting $\mathbf{Z}_i^{(R)}$ into eq. (8). Repeating the above procedure section by section, the given radiation impedance matrix $\mathbf{Z}_{rad}$ in eq. (6) is transformed into the input impedance matrix of the first section.

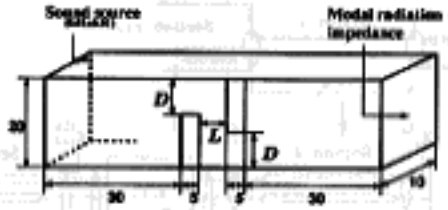
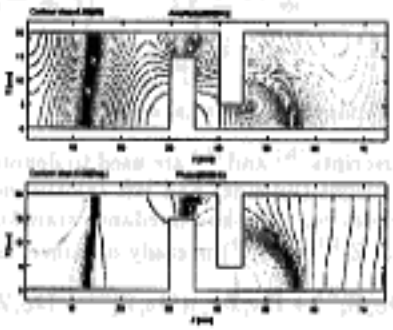
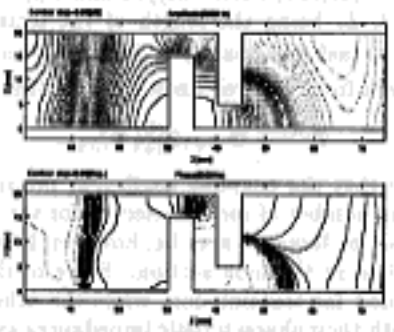


Figure 2: 5-section configuration.



(a) Proposed method



(b) TLM simulation

Figure 3: Sound-pressure distributions at 6 kHz.

Once all input impedance matrices are obtained, wave component vectors a and b at each section can be calculated using eqs. (3) and (4) recursively. Then sound-pressure distribution is obtained from eq. (1). A transfer function H of the proposed model is evaluated by,

$$H \propto \frac{\sqrt{W_t}}{U_G} \tag{14}$$

where W_t and U_G are a radiation power and a source volume velocity, respectively.

3. COMPUTATIONAL RESULTS

3.1. Configuration with 5-section

A 5-section configuration is used to imitate the occlusion at the teeth as illustrated in figure 2. A tiny square sound source (0.01 mm × 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 4 higher-order modes (1,0),(0,1),(2,0) and (0,2) are considered for the mode coupling at each junction. Computed sound-pressure distribution at 6 kHz for the size of $D = L = 5.0$ mm is shown in figure 3(a), both amplitude and phase contours being presented. Note that all higher-order modes are evanescent at 6 kHz. Figure 3(b) shows the result obtained from the TLM simulation with an element length of 1.0 mm. These results confirm that the proposed model can properly represent the acoustic characteristics of 3-D tube configuration.

The superposition of the plane waves and the 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 1-D vocal tract model, such as estimating an equivalent length L_e and area S_e of the occlusive section for the plane waves to propagate in the vertical direction as illustrated in figure 4. 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:

$$\begin{bmatrix} p^{(R)}_{2_{08}} \\ S_2 V^{(R)}_{2_{80}} \end{bmatrix} = \begin{bmatrix} \cos(kL_e) & j\frac{\rho c}{S_e}\sin(kL_e) \\ j\frac{S_e}{\rho c}\sin(kL_e) & \cos(kL_e) \end{bmatrix} \begin{bmatrix} p^{(L)}_{4_{00}} \\ S_4 V^{(L)}_{4_{00}} \end{bmatrix} \tag{15}$$

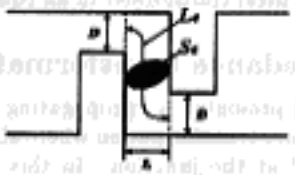
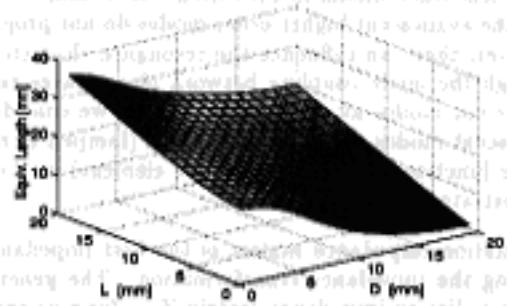
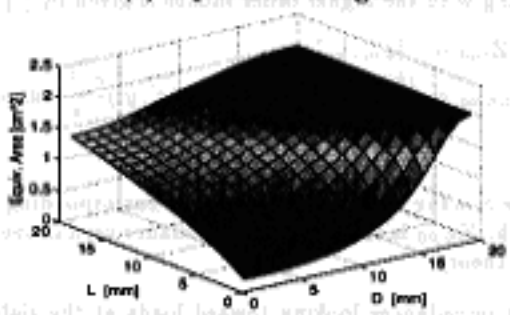


Figure 4: Equivalent length L_e and area S_e.



(a) Equivalent Length



(b) Equivalent Area

Figure 5: Equivalent length and area at the occlusion as a function of D and L.

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. Figure 5 shows L_e and S_e as a function of D and L. As imagined from figure 4, the increase of L_e and the decrease of S_e can be seen as D becomes shorter. Note that L_e and S_e coincide with the geometrical values when there exists no occlusion ($D = 20$ mm). 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 3-D 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 1-D model.

3.2. Configuration based on MRI data

The 3-D shape of a /ʃ/ measured by MRI has been converted into 36-section configurations to show the influence of the offset connection of tubes upon the transfer functions. First each tube is aligned to a common horizontal plane with keeping the symmetry with respect to the lateral direction. Then the axis position of each tube is perturbed randomly both in the vertical (z) and lateral (y) directions. The maximum perturbation is 5 percent of cross-sectional size. Figure 6 shows an example of the geometry of tubes. The resultant tube configuration is slightly asymmetrical with respect to the lateral direction. The transfer functions for 20 configurations are shown in figure 7. 5 higher-order modes (3 in lateral and 2 in vertical direction) are considered for the computation. Higher formant frequencies above 5 kHz are largely influenced by the small change of the axis position. Note that there is no difference in area functions for these configurations. The appearance of zeros and their frequencies tend to be more sensitive to the small change of axis position.

4. CONCLUSION

It has been shown that the proposed model is effective to analyze the acoustic characteristics of 3-D configurations of rectangular tubes used as a geometrical approximation of vocal tracts. In particular, the proposed method presents the following advantages: (1) more accurate acoustic characteristics, compared to those obtained from the 1-D modeling; (2) shorter computational time compared to FEM and/or TLM. These are useful features for the analysis of the vocal tract acoustics and/or the improvement of speech synthesis systems based on vocal tract models.

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 Prof. N.Miki, Future University-Hakodate, for his valuable suggestions for this work.

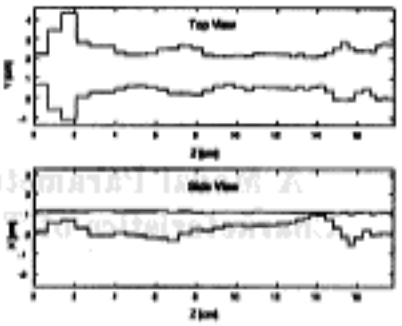


Figure 6: Example of 36-section configuration.

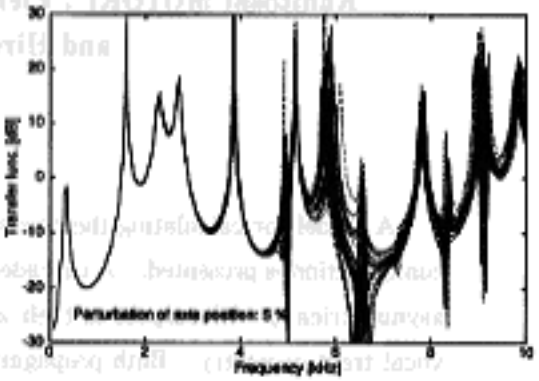


Figure 7: Transfer functions with 5 percent perturbation of axis position.

Part of this work has been supported by a research project of High-Tech Research Center, Hokkai-Gakuen University, 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).