Coupling Parameters Estimation for Two-layer Model of the Velum

Xuefeng Zhou, Xugang Lu*, Jianwu Dang*, and Pierre Badin***

*Acoustic Information Lab, School of Information Science
Japan Advanced Institute of Science and Technology, Ishikawa, Japan 923-1292
**ICP CNRS UMR 5009 & INPG & University Stendhal, Grenoble, France
{s-shuu, xugang, jdang }@jaist.ac.jp

Abstract

The velum has an important function in coupling the nasal to the oral cavities during producing a nasal sound. In this paper, the velum is treated as a mechanical model with two layers, and is further approximated an equivalent electricity circuit to combine with the acoustical model of the vocal tract. To estimate the parameters for the model, the sound pressures in the intraoral and intranasal and the vibration of the velum were measured simultaneously. A minimizing square error optimization method was applied on the measurements to estimate parameters of two-port circuit. For the convenience of calculation, we suppose that there is only phase delay and amplitude difference between volume velocity of the intra-oral and volume velocity of the intra-nasal. By deducing the estimation formulas, we could get analytical linear functions that could be solved easily. These estimated parameters could explain the observed measurements data well. Based on these parameters, further refinement of the real parameters can be carried out.

1. Introduction

To improve the quality of sounds synthesized based on speech production model, modeling of the coupling between the oral cavity and the nasal cavity is an important issue. In order to complete a physiological model of speech production based on the movements of speech organs, it is necessary to determine how the velum interacts with the other speech organs during speech production, and what effect is observed on characteristics of speech sound when the movements of the velum change the coupling degree between the oral and nasal cavities. Moll and Daniloff used high-speed cinefluorographic films to investigate the timing of velar movements during speech for four normal subjects [1]. They measured velar movement and estimated velopharyngeal opening. In the previous studies [2], speech sound was separated into three speech signals radiated from the lips, the nostrils and the pharynx wall. The results showed that a large nasal sound radiated during production of closed vowels and voiced stop consonants, where the velum had been expected to be closed. It implied that the sound radiation from the nostrils might be caused by velum vibration. To improve the acoustic model for vowel production, Dang and Honda estimated the acoustic characteristics of the velum based on measurements using acoustic and mechanical methods [4]. Their results showed that the vowel quality was improved by taking the transvelar effects into the acoustical model. The purpose of this research is to improve vocal tract model for production of vowels and voiced stop consonants by considering the transvelar nasal coupling via the velum vibration. For this purpose, we investigate the mechanical and acoustical behaviors of the velum and refine the velum model.

2. Consideration of the velum model

2.1 Description using a two-layer plate

In coupling of the nasal and oral cavities, the velum works not only as a valve, which open or close the velopharyngeal port, but also couples two cavities by its vibration when the velopharyngeal port is closed. In mechanical point of view, the yielding velum can be considered as a two-layer plate. Figure 1 (a) shows the configuration of the mechanical model for the velum. From the mechanical model, we can get an equivalent electrical circuit, which shown in Figure 1 (b).

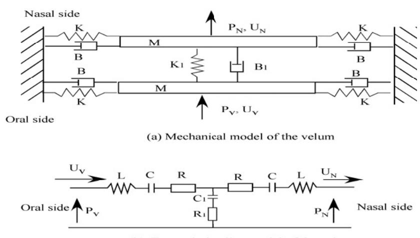

Fig.1 Modeling the velum with electricity circuit

2.2 Description in electric circuit

For the two-layer model, it can be regarded as two-port (one-input and one-output) circuit with some inner parameters. For simplification of the deduction, we use a simplified Fig.2 as the velum model

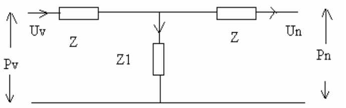

Fig.2 Simplified velum model

In frequency domain, the impedance can be written as:

$$Z(\omega)=R+j\omega L+\frac{1}{j\omega C}=R+j(\omega L-\frac{1}{\omega C})$$
$$Z_1(\omega)=R_1+\frac{1}{j\omega C_1}=R_1+j(\frac{-1}{\omega C_1}) \qquad (1)$$

For convenience of calculation, we suppose $K=\frac{1}{C}, K_1=\frac{1}{C_1}$, then formula (1) can be changed into:

$$Z(\omega)=R+j\omega L+\frac{1}{j\omega C}=R+j(\omega L-\frac{K}{\omega})$$
$$Z_1(\omega)=R_1+\frac{1}{j\omega C_1}=R_1+j(\frac{-K_1}{\omega}) \qquad (2)$$

Based on circuit analysis, we can get:

$$\begin{bmatrix} P_v \\ P_n \end{bmatrix}=\begin{bmatrix} Z+Z & -Z_1 \\ Z_1 & -(Z+Z_1) \end{bmatrix}\begin{bmatrix} U_v \\ U_n \end{bmatrix} \qquad (3)$$

It is obvious that if the P_v, P_n, U_v, U_n can be measured and got, the impedance parameters will be gotten easily. But in real measurement for the two-port circuit model, only some necessary variables can be gotten, and another one variable for volume velocity in intra-oral side can not be measured. The measurement is done in next part.

3. Input and output of the two-port

To estimate the parameters of the two-port circuit, this study employed the previous measurements in [4]. Here, we give a brief explanation for the measurements. Figure 3 shows the setup for the measurement. M1is the microphone in front of the lips; M2 is a probe microphone inside the oral cavity; M3 is a probe microphone inside the nasal cavity; A1 is an accelerometer on the velum in the nasal side; A2 is an accelerometer on the wall of the nostrils.

There are three measurements concerned with our estimation: acceleration from A1, intraoral sound pressure from M2, and intranasal sound pressure from M3. The acceleration can be used to calculate the volume velocity of intra-nasal by integrating them over time. For the two-layer model of velum, the oral sound pressure was considered to be the input to the velum, and the nasal sound pressure and the volume velocity of the intra-nasal were the outputs. We could not measure the volume velocity on the oral side, which is required for the estimation.

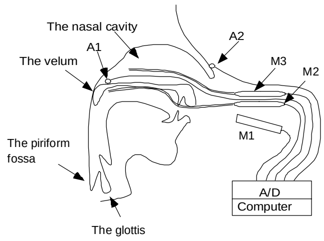

Fig.3 the setup for measuring sound pressures and accelerations

The microphone outputs of M2 and M3 are regarded as the parameters P_v, P_n of the equation (3). for the velum. The volume velocity (U_n) in the nasal side can be calculated by integrating the output of the acceleration A_1 over time. The area of the velum S=4cm^2; dt=1/12000 (sec); (Sampling frequency is 12kHz)

$$U_n = S\int A_1(t)dt \qquad (4)$$

For convenience of calculation, all calculations are done in frequency domain, the corresponding volume velocity calculation in frequency domain as:

$$U_n(w)=\frac{S*A_1(w)}{jw} \qquad (5)$$

Since U_v can not be gotten easily, the left problem is to estimate the model parameters based on the measurement of P_v, P_n, U_n.

4. Optimization for model parameters

After getting $P_v, P_n\ U_n$, we can use optimization algorithm to estimate the model parameters. Even this model is simple, the optimization will bring nonlinear function, and analytical solution cannot be gotten. For further convenience of calculation, we suppose that there is only phase delay and amplitude difference between U_v and U_n. We can write them in Fourier domain as:

$$U_v(w) = kU_n(w)e^{jwt_0} \quad (6)$$

so :

$$U_v(w) = |U_v(w)|\exp(\theta_v(w))$$

$$U_n(w) = |U_n(w)|\exp(\theta_n(w)) \quad (7)$$

So we can get:

$$k = \frac{|U_v(w)|}{|U_n(w)|} \quad (8)$$

$$t_0 = \frac{\theta_v(w) - \theta_n(w)}{w} \quad (9)$$

$$\theta_v(w) = \arg(U_v(w)) = arctg(\frac{imag(U_v(w))}{real(U_v(w))})$$

$$\theta_n(w) = \arg(U_n(w)) = arctg(\frac{imag(U_n(w))}{real(U_n(w))}) \quad (10)$$

It is obvious that the parameters k, t_0 are variables with frequency. But usually, when uttering a sound, the velum parameters are almost fixed in the stable stage, so the k, t_0 is fixed and can be estimated by observing the phase delay and amplitude difference of the air pressure of the two intra sides of velum. Based on formula (3) and (6), we can get

$$P_n = P_v - Z(U_v + U_n)$$
$$= P_v - Z(kU_n e^{jwt_0} + U_n) \quad (11)$$

Formula (11) can be changed into:

$$Z = \frac{P_v - P_n}{U_n(1 + k\exp(jwt_0))} \quad (12)$$

Because the frequency response of the impedance has only one minimum value, the corresponding optimization algorithm should find parameters of k, t_0 for matching this character. Apparently, the left variable has one minimum value in amplitude when

$$w_i = \frac{1}{\sqrt{LC}} \quad (13)$$

For further calculation, formula (12) is changed into:

$$|Z(1 + k\exp(jwt_0))| = |\frac{P_v - P_n}{U_n}| \quad (14)$$

Suppose

$$|f_r(w)| = |\frac{P_v - P_n}{U_n}| \quad (15)$$

this value can be calculated by real measurements. The theory prediction for this value is from:

$$f(w) = Z(1 + k\exp(jwt_0)) \quad (16)$$

For convenience of optimization calculation, we define the objective function as:

$$F(R,L,C,k,t_0) = \sum_w [E(w) \times E(w)^*] = \sum_w |E(w)|^2 \quad (17)$$

Where $E(w) = f_r^2(w) - f^2(w)$, $E(w)^*$ is complex conjugate of $E(w)$.

For convenience of optimization, the k, t_0 is fixed, and can be estimated before this optimization. Based on this assumption, the optimization can be done as:

$$\frac{\partial F}{\partial R} = 0 ; \frac{\partial F}{\partial L} = 0 ; \frac{\partial F}{\partial C} = 0 \quad (18)$$

By solving this optimization equation, we can get a linear function as:

$$AX = B \quad (19)$$

Where X is a vector formed by model parameters. The model parameters will be solved based on this equation.

Since the vibration of velum might bring some poles or zeros in the transfer function (it is obvious that only one pole is introduced in this two-layer model), the optimization should be done only one the frequency points as in formula (13). This point can also be regarded as a constraint for the optimization. Analyzed from formula (12), (14), we should estimate the parameters by only focusing on those lower valleys of the profile of the formula (15) (this will be discussed in next part). The profile can be extracted via cepstral analysis by keeping only lower order cepstral coefficients, and reconstructed it into frequency domain to get the spectral profile.

5. Model parameters estimation

Based on formula (14), the left side of the equation is the theoretic result based on velum model, the right side is the result from real measurement. First, if the model parameters are known, let us see what the theoretic simulation for left side of equation (14). For given conditions: R1=100; R=6.25; L=1.875e-2; C=9.47e-5; C1=1.99e-5; k=1.5, t0=0.325ms

The simulation result is shown in Fig.4

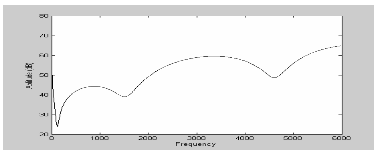

Fig.4 Theoretic simulation for left side of formula (14)

It is obvious that the optimization should be done to match the lower profile valley with real measurement. Based on the above analysis, we estimated the model parameters. Databases are recorded from three persons. First, the recording database is segmented with starting and ending points manually. The calculation is in frequency domain, the Fourier transform is averaged frame by frame, the FFT length is 1024 points with a hamming window, and frame shift is 256 points. After getting the averaged Fourier transform, formula (13) is calculated, and the result is smoothed with cepstral analysis to get the profile. In this paper, the voice bar of the voiced stops in /ba/, /be/, /bi/, /bu/, /bo/ is selected. The calculation results are shown in table 1:

t0	0.325ms
K	1.5
R	52.1919g/s/cm^4
L	10.0262g/cm^4
C	2.8081e-005s^2cm^4/g

Table 1 Estimation results

During the optimization, a lot of assumptions are used to form an analytical linear equation, the assumptions will bring some errors for this estimation. This estimation is only an initial estimation for the parameters, more accurate estimation will be designed to get more faithful parameters for our further investigation.

6. Discussion

Velum vibration is the factor to change the spectral structure for nasal sounds, usually the primary frequency range for nasal sounds is between 200 and 2500Hz [5]. Nasals are characterized by a stable concentration of energy in the lower frequency regions with a first formant at around 300 Hz, so in above optimization algorithm, only lower frequency band (below 1kHz) is used for optimization. For co-articulation mechanism, the nasal will influence the lower formants structure of the following vowels. By this investigation, we can model the mechanism for articulation, and find out the place for co-articulation. It will be helpful for speech synthesis and speech recognition. Further accurate investigation should be done.

7. References

[1] Moll, K. L., Shriner, T. H. “Preliminary investigation of a new concept of velar activity during speech.” Cleft Palate J., 4,58-69. (1967)

[2] Suzuki, H., Dang, J., and Nakai, K. “Measurement of sound vibration at the lips, nostrils and pharynx wall in speech utterance and simulation of sound leakage from the oral cavity to the nasal cavity in non-nasal sound.” Jpn. IEICE Trans., J74-A,1705-1714 (in Japanese) . (1991)

[3] Dang, J., Nakai, K., Suzuki, H. Measurement and simulation of intraoral pressure and radiation of stop consonant. J. Acoust. Soc. Jpn., 59, 313-320. (in Japanese). (1993)

[4] Dang, J., Honda, K. "An improved vocal tract model of vowel production implementing piriform fossa resonance and transvelar nasal coupling," ICSLP96, 965-968 (1996)

[5] Kenneth N. Stevens, Acoustic Phonetics, MIT press, Cambridge, Massachisetts, London, England, 1998