OBTAINING THE VOCAL-TRACT AREA FUNCTION FROM THE VOWEL SOUND

Huiqun Deng, Michael P. Beddoes, Rabab K. Ward, Murray Hodgson*
Electrical and Computer Engineering Department, *Mechanical Engineering Department
University of British Columbia, 2356 Main Mall, Vancouver, BC, Canada V6T 1Z4
huid@ece.ubc.ca, mikeb@ece.ubc.ca, rababw@icics.ubc.ca, hodgson@mech.ubc.ca

1. INTRODUCTION

Vocal-tract area functions (VTAFs) are needed in speech synthesis, speech recognition, and the detection of the vocal-tract shape. The vocal-tract area function can be measured using X-ray or MRI methods. But, both methods are time-consuming and not convenient. It has long been desired to obtain the vocal-tract area function from the speech signal.

It is shown that the VTAF can be derived from the vocal-tract filter (VTF) assuming the wall of the vocal tract is rigid, the length of the vocal tract is known, and the lip or the glottal reflection coefficient is 1 [1,2]. In Atal's method for deriving the VTAF from the VTF, the vocal tract is assumed to completely close at the glottis, and to be terminated with characteristic impedance at the lip opening [1]. In Wakita's method, the vocal tract is assumed to be terminated with characteristic acoustic impedance at the glottal end, and with zero acoustic impedance at the lip opening [2]. However, these assumptions about the boundary conditions cannot be satisfied all the times. The glottal reflection coefficient is time varying, because the glottis opens and closes periodically during voicing. The lip radiation impedance can only be approximated as zero at low frequencies, and can be characteristic impedance only when the lip opening is connected with a reflectionless tube.

Accurate estimation of the VTAF requires the VTF estimated from a vowel signal should not contain the influence of the glottal wave, and the influence of the non-ideal glottal and lip boundary conditions. The method for eliminating the influence of the glottal wave on the VTF estimation from a vowel sound signal is developed in [3]. In this paper, we investigate the effect of non-ideal glottal and lip boundary conditions on the estimation of the VTAF.

2. THE VOCAL-TRACT FILTER

The acoustic effect of the vocal tract can be modeled using a multi-sectional cylindrical tube, with each section having the same length and different cross-sectional area (Fig.1). The signal flow diagram from the glottal wave U_g to the lip volume velocity U_{lip} can be represented in terms reflection coefficient r_i and the delay D in each section, r_g (the glottal reflection coefficient), and r_{lip} (the lip

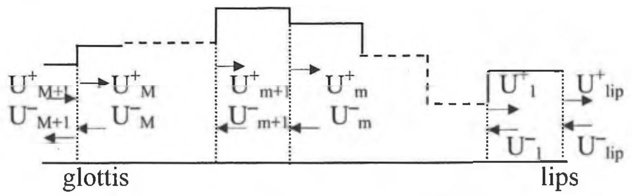

Fig. 1. The acoustic tube model of the vocal tract.

reflection coefficient), as shown in Fig. 2 [3], where r_{lip} is:

$$r_{lip} = (\rho c / A_1 - Z_{lip})/(\rho c / A_1 + Z_{lip}) \quad (1)$$

where Z_{lip} is the lip radiation impedance, A_1 is the cross-sectional area of section 1, and ρc is the characteristic impedance of the air. r_g is defined as:

$$r_g = (Z_g - \rho c / A_M)/(Z_g + \rho c / A_M) \quad (2)$$

where Z_g is the acoustic glottal impedance, A_M is the cross-sectional area of section M,

The transfer function from the glottal wave to the lip volume velocity is an all-pole filter, which is denoted as TF_{GL}. It is time varying due to the time varying r_g. The all-pole filter estimated from speech signals using LPC [1] is usually called as vocal-tract filter. But, it is actually an averaged version of the TF_{GL}, and is some different from what is required in the estimation of the VTAF. In Atal's method, the VTF used for estimating the VTAF is defined to be the complex ratio of the total volume velocity at the lips to the total volume velocity at the backend of the vocal tract, i.e., $U_{lip}/(U^+_M+U^-_M)$. The TF_{GL} is identical to the VTF in Atal's method, only if rg=1. In Wakita's method, the VTF is defined to be the complex ratio of the total volume velocity at the lips to the volume velocity entering the glottis from the trachea, i.e., U_{lip}/U^+_{M+1}, assuming $Z_{lip}=0$, i.e., $r_{lip}=1$. The TF_{GL} is identical to the VTF in Wakita's method, only if $r_{lip}=1$.

3. VOCAL-TRACT BOUNDARY CONDITIONS AND VOCAL-TRACT AREA FUNCTION ESTIMATION

In order to see the effect of non-ideal r_g and r_{lip} contained in the TF_{GL} on the estimation of the VTAF, we synthesize the TF_{GL} for a given VTAF, and different r_{lip}'s and r_g's, and use the synthetic TF_{GL} to estimate the VTAF. The difference between the estimate and the given VTAF is

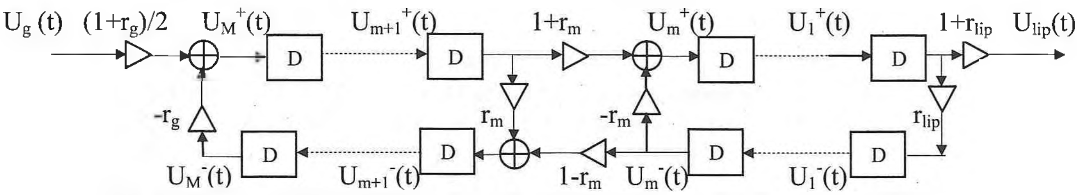

Fig. 2. The signal flow diagram from the glottal wave to the lip volume velocity.

the effect of the non-ideal r_g and r_{lip}. The VTAF of /ɑ/ measured using the magnetic resonance imaging method [4] is used in synthesizing the impulse response of the TF_{GL}. From the synthetic impulse response, we estimate the VTAF under different assumptions about the vocal-tract boundary conditions. If the vocal-tract boundary conditions are assumed to be r_g=1 and r_{lip} is arbitrary, the vocal-tract reflection coefficients are estimated using method 1:

$$r_m = - k_{m-1} \qquad m=1,2,\ldots\ldots,M-1 \qquad \text{(method 1)}$$

where $r_m=(S_m-S_{m+1})/(S_m+S_{m+1})$, S_m is the cross-sectional area of section m, m increases from the glottis to the lips, k_m is defined in Eq. (11) in [2]. This method is derived through matching Atal's acoustic filtering process [1] to Wakita's mathematical filtering process of the VTF [2]. If the vocal-tract boundary conditions are assumed to be r_{lip}=1 and r_g is arbitrary, the vocal-tract reflection coefficients are estimated using method 2:

$$r_m = k_{m-1} \qquad m=1,2,\ldots\ldots,M-1 \qquad \text{(method 2 [2])}$$

where m increases from the lips to the glottis [2]. Let S_1 =1, the normalized cross-sectional areas, which form the VTAF, are then obtained:

$$S_{m+1}=S_m(1- r_m)/(1+ r_m) \qquad m=1,2,\ldots\ldots, M-1 \qquad (3)$$

The frequency responses of the synthetic TF_{GL}'s and the VTAF estimates derived from the synthetic TF_{GL}'s using methods 1 and 2 for /ɑ/ are shown in Fig. 3. The bandwidths of the resonance of the VTF are damped, which represent the energy loss in the TG_{GL}, if r_g or r_{lip} is small. The influence of the glottal loss and lip loss on the estimation of the VTAF can be seen comparing the estimates with the given VTAF used for synthesizing the VTF. Our results show that the VTAF can be recovered from the TF_{GL} using method 1, if the TF_{GL} corresponds to r_g=1; or, using method 2, if the TF_{GL} corresponds to r_{lip}=1. Method 1 works well only if the glottal reflection coefficient is one, not being affected by different lip reflections. Method 2 works well only if the lip reflection coefficient is one, not being affected by different glottal reflections.

4. DISCUSSION

Although method 2 is not sensitive to r_g, it is sensitive to r_{lip}. r_{lip}=1 (i.e. Z_{lip}=0) is true for ka<<1, where $k=2\pi f/c$, a is the lip opening radius. Therefore, only low sampling rate (F_s< 7 kHz) should be used in method 2. The sampling rate F_s determines the number of sections of the tube model: $M=2LF_s/c$, where L is the length of the vocal tract, and c is the sound speed. Thus, method 2 cannot obtain detailed structures of the VTAF.

Method 1 is not subject to r_{lip}. Therefore, it can work well over wide frequency range, and can allow high sampling rate. Thus, method 1 can obtain more detailed structures of the VTAF than method 2. For more accurate estimation of the VTAF, the TF_{GL} used in method 1 should be estimated from the speech signal recorded during closed phases of the glottis, and the speech signal should be recorded in a reflectionless tube connected to the lip opening.

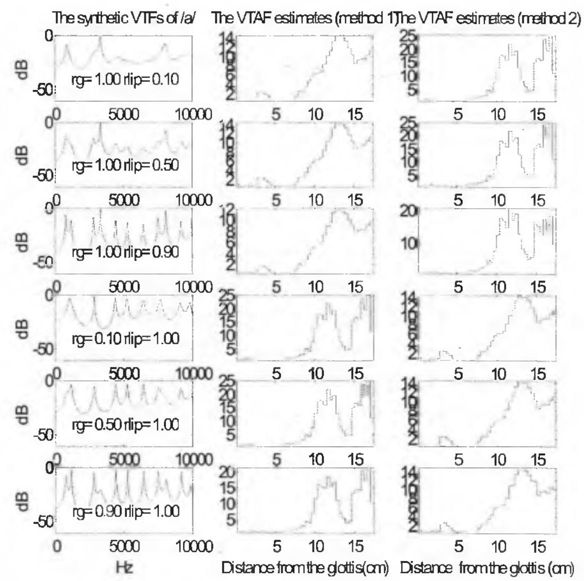

Fig. 3. The synthetic TF_{GL}'s and the estimates of the VTAF.

REFERENCES

[1] Atal, B. S. and L. Hanauer, "Speech Analysis and Synthesis by Linear Prediction of the Speech Wave", *J. Acoust. Soc. Amer., Vol. 50, Number 2 (part 2), 1971, p 637-655.*

[2] Wakita, H., "Direct Estimation of the Vocal Tract Shape by Inverse Filtering of Acoustic Speech Waveforms", *IEEE Trans. Audio Electroacoust. Vol.AU-21,*1973, p *417-427.*

[3] Deng, H., M. P. Beddoes, R. K. Ward and M. Hodgson, "Estimating the Derivative of the Glottal Wave and the Vocal-tract Filter from a Vowel Sound Signal", IEEE PACRIM'03 Conference on Communications Computers, and Signal Processing, Aug. 28-30, 2003, Victoria, Canada.

[4] Story, B. H. and R. Titze, "Vocal tract area functions from magnetic resonance imaging", *J. Acoust. Soc. Amer., Vol. 100,1995, p 537-554.*