EFFECTS OF GLOTTAL AND LIP BOUNDARY CONDITIONS ON VOCAL-TRACT AREA FUNCTION ESTIMATES FROM SPEECH SIGNALS

Huiqun Deng[1], Rabab K. Ward[1], Michael P. Beddoes[1], Murray Hodgson[2]

[1]Electrical and Computer Engineering Department, [2]Mechanical Engineering Department
The University of British Columbia, Vancouver, BC V6T 1Z4, Canada
huid@ece.ubc.ca, rababw@icics.ubc.ca, mikeb@ece.ubc.ca, hodgson@mech.ubc.ca

ABSTRACT

High-resolution vocal-tract area functions (VTAFs) can be derived from vocal-tract filters (VTFs) estimated from vowel sound signals with a wide bandwidth. However, the effects of open glottises and frequency-dependent lip reflection coefficients contained in the VTF estimates distort the VTAF estimates. Given VTF estimates obtained over closed glottal phases, we provide a method for eliminating the distortion effects of frequency-dependent lip reflection coefficients on the VTAF estimates. When the VTF estimates contain limited effects of incomplete glottal closures, this method can still obtain reasonable VTAF estimates if the vowel sounds are produced with large lip openings. The VTAF estimates obtained using our method from sounds /ɑ/ produced by different subjects are very similar to that measured using the magnetic resonate imaging method. Theoretically, to eliminate both distortions caused by lip reflection coefficients and incomplete glottal closures in the VTAF estimates, lip-opening areas must be known.

1. INTRODUCTION

Obtaining vocal-tract area functions (VTAFs) from speech signals has applications in speech synthesis, speech recognition, speech pathology, language training, etc. It is shown that a unique vocal-tract area function can be derived from a vowel sound under one of two special boundary conditions [1] [2]. Boundary condition 1: the lip opening is terminated with zero impedance (i.e., the lip reflection coefficient r_{lip} is one); the glottal end is terminated with some characteristic impedance [2]. Boundary condition 2: the lip opening is terminated with some characteristic impedance; the glottal end is completely closed (i.e., the glottal reflection coefficient r_g is one) [1]. In reality, however, these boundary conditions can not always be true: the glottal area is time-varying during phonation; the lip radiation impedance varies with frequency. Consequently, VTAF estimates obtained assuming boundary condition 1 or 2 are distorted.

To avoid the effects of open glottises and glottal waves on VTAF estimates, a method for obtaining VTAF over closed glottal phases has been developed [3]. The remaining problem is to eliminate the distortions caused by frequency-dependent lip boundary conditions. Although one can obtain VTAF estimates from signals in a low frequency range (0-4 kHz), over which boundary condition 1 is nearly satisfied, such speech signals lack information about VTFs over the higher frequency range, and the resulting VTAF estimates lack details of vocal-tract shapes. To obtain high-resolution estimates of VTAFs, speech signals with a wide bandwidth must be used. In these cases, the distortion effects of frequency-dependent lip boundary conditions on VTAF estimates need to be eliminated. Given a VTF estimate obtained over closed glottal phases, this paper presents a method for eliminating the distortion effect of the frequency-dependent lip reflection coefficient on the VTAF estimate. In the following, the concepts related to VTF estimates are first clarified. Next, the model of the frequency-dependent lip reflection coefficient in the VTF, and the method for eliminating the effect of the frequency-dependent lip boundary condition on the VTAF estimate are developed. Finally, the simulation and experimental results are presented.

2. FROM SPEECH SIGNALS TO VOCAL-TRACT AREA FUNCTIONS

2.1 Transfer functions for producing a vowel sound

The acoustic system for producing a vowel sound is modeled as shown in Fig. 1 [4], where $u_{sc}(t)$ is the equivalent glottal volume velocity source and Z_g is the glottal impedance; $u_g(t)$ and $p_1(t)$ are the total volume velocity and the sound pressure at the back end of the vocal tract, respectively; Z_{lip} is the lip radiation impedance; $u_{lip}(t)$ and $p_{lip}(t)$ are the total volume velocity and sound pressure at the lip opening, respectively.

The transfer function of a vocal-tract filter (VTF) is defined as:

$$H_{VTF}(f)=U_{lip}(f)/U_g(f) \tag{1}$$

where $U_g(f)$ and $U_{lip}(f)$ are the Fourier transforms of $u_g(t)$ and $u_{lip}(t)$, respectively. The transfer function of a glottal-vocal-tract filter (GVTF) is defined as:

$$H_{GVTF}(f)=U_{lip}(f)/U_{sc}(f) \tag{2}$$

where $U_{sc}(f)$ is the Fourier transform of $u_{sc}(t)$. H_{GVTF} contains the effect of non-linear and time-varying glottal impedance, and thus is non-linear and time varying. As seen in Fig. 1, H_{GVTF} equals H_{VTF} when the glottis is closed, and thus an H_{VTF} estimate can be obtained over closed glottal phases.

The vocal tract is modeled as an acoustic tube with M equal-length sections [1, 2], as shown in Fig. 2, where $u_1^+(t)$ and $u_1^-(t)$ are the positive-going and negative-going volume velocities at the left end of section 1. $u_1^+(t)+u_1^-(t)=u_g(t)$ (note: in this study, the reference directions of positive-going and negative-going volume velocities are the same). If the sampling rate of the signal is F_s=0.5Mc/L, where L is the length of the vocal tract, and c is the sound speed, then H_{GVTF} in the Z domain is [5]:

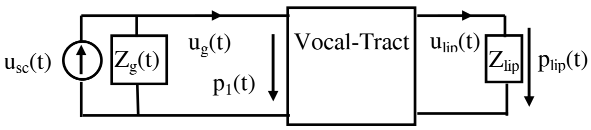

Fig. 1. The acoustic system for producing vowel sounds.

$$H_{GVTF}(z) \equiv \frac{U_{lip}(z)}{U_{sc}(z)} = \frac{0.5z^{-M/2}(1+r_g)(1+r_{lip})\prod_{m=1}^{M-1}(1+r_m)}{[1, r_g]\begin{bmatrix}1 & r_1\\ r_1 z^{-1} & z^{-1}\end{bmatrix}\cdots\begin{bmatrix}1 & r_{M-1}\\ r_{M-1}z^{-1} & z^{-1}\end{bmatrix}\begin{bmatrix}1 & r_{lip}\\ r_{lip}z^{-1} & z^{-1}\end{bmatrix}\begin{bmatrix}1\\ 0\end{bmatrix}} \quad (3)$$

where:

$$r_g = (Z_g - \rho c / S_1)/(Z_g + \rho c / S_1) \quad (4)$$

$$r_m = (S_{m+1} - S_m)/(S_{m+1} + S_m) \quad (5)$$

$$r_{lip} = (\rho c / S_M - Z_{lip})/(\rho c / S_M + Z_{lip}) \quad (6)$$

S_m is the m[th] cross-sectional area and ρ is the density of air. Denote:

$$\begin{bmatrix}A_m(z) & B_m(z)\\ C_m(z) & D_m(z)\end{bmatrix} = \begin{bmatrix}1 & r_1\\ r_1 z^{-1} & z^{-1}\end{bmatrix}\cdots\begin{bmatrix}1 & r_m\\ r_m z^{-1} & z^{-1}\end{bmatrix} \qquad m=1,\ldots,M \quad (7)$$

then:

$$H_{GVTF}(z) = \frac{0.5z^{-M/2}(1+r_g)(1+r_{lip})\prod_{m=1}^{M-1}(1+r_m)}{A_{M-1} + r_g C_{M-1} + (B_{M-1} + r_g D_{M-1}) r_{lip} z^{-1}} \quad (8)$$

The VTF transfer function can be obtained from H_{GVTF} with $r_g=1$ (since $Z_g=\infty$ when the glottis is closed) in Eq. (8):

$$H_{VTF}(z) = \frac{z^{-M/2}(1+r_{lip})\prod_{m=1}^{M-1}(1+r_m)}{A_{M-1} + C_{M-1} + (B_{M-1} + D_{M-1}) r_{lip} z^{-1}} \quad (9)$$

2.2 VTAF estimation based on boundary condition 2

It can be shown that for m=1, …, M, $A_m(z)+C_m(z)$ is an m[th] order polynomial in z^{-1} with leading term of unity, and that $B_m(z)+D_m(z)$ is the reciprocal polynomial of $A_m(z)+C_m(z)$ [1]. Denote:

$$A_m(z) + C_m(z) = 1 + g_1 z^{-1} + \ldots\ldots + g_m z^{-m} = G_m(z) \quad (10)$$

then:

$$B_m(z) + D_m(z) = g_m + g_{m-1} z^{-(m-1)} + \ldots\ldots + z^{-m} = z^{-m} G_m(z^{-1}) \quad (11)$$

It is known that $G_{m-1}(z)$ can be derived from $G_m(z)$ using:

$$G_{m-1}(z) = (G_m(z) - r_m z^{-m} G_m(z^{-1}))/(1 - r_m^2) \quad (12)$$

It can be shown that the coefficient of z^{-m} in the polynomial $G_m(z)$ equals r_m [1]. If $r_{lip}=r_M$=constant, the denominator of $H_{VTF}(z)$ becomes $G_M(z)$ and the coefficient of z^{-M} in $G_M(z)$ is r_M. Then, r_{M-1} can be derived from $G_M(z)$ using Eq. (12). Similarly, other r_m's can be derived. Then, the relative areas of the VTAF can be derived from r_m's using Eq. (5).

As mentioned earlier, if the $H_{VTF}(z)$ estimates cover a wide frequency range, over which r_{lip} is frequency-dependent, then r_m's derived based on the assumption that r_{lip}=constant become distorted. In the next section, to eliminate the distortion effect of the frequency-dependent r_{lip} from the estimates of r_m's, we model the effect of the lip reflection coefficient in the VTF.

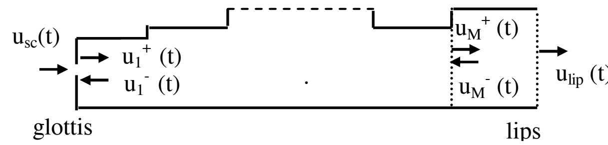

Fig. 2. The acoustic tube model of the vocal tract

2.3 Modeling the lip reflection coefficient

We represent the lip radiation impedance at very low frequencies, using that of an unflanged pipe with the same opening area as the lip opening, and at very high frequencies, using that of a piston with the same opening area as the lip opening in an infinite baffle. Since the radiation impedance of an unflanged pipe is approximately half that of the piston in an infinite baffle [6] and since the reflection effect of a subject's head and body gradually increases as frequency increases, we represent the normalized Z_{lip} at frequency f as:

$$Z_{lip} S_M / \rho c = (0.5 + f / F_s) Z_p \qquad 0 < f < F_s / 2 \quad (13)$$

where S_M is the lip opening area and Z_p is the normalized radiation impedance of the piston in an infinite baffle. It can be shown that r_{lip} is a low-pass filter. In this study, a first-order IIR filter is used to represent r_{lip} in the Z domain:

$$r_{lip}(z) = \mu(1 + \beta z^{-1})/(1 + \alpha z^{-1}) \quad (14)$$

where α and β depend on lip opening area, $\mu = (1+\alpha)/(1+\beta)$, and $r_{lip}(z=1)=1$, since z=1 corresponds to zero frequency.

2.4 Eliminating the effect of r_{lip} on the VTAF estimate

Substituting Eq. (14) for r_{lip} in Eq.(9), the VTF transfer function becomes:

$$H_{VTF}(z) = \frac{[(1+\mu+(a+\beta\mu)z^{-1}]z^{-M/2}\prod_{m=1}^{M-1}(1+r_m)}{[A_{M-1}(z)+C_{M-1}(z)](1+az^{-1}) + [B_{M-1}(z)+D_{M-1}(z)]z^{-1}(1+\beta z^{-1})\mu} \quad (15)$$

The denominator of the $H_{VTF}(z)$ in Eq. (15) can be expressed in terms of G_{M-1} using Eqs. (10) and (11). Also, it can be estimated from a speech signal corresponding to closed glottal phases using the method in [3]. Let the estimate of the denominator of $H_{VTF}(z)$ be $1+h_1 z^{-1}+\ldots\ldots+h_{M+1}z^{-(M+1)}$. Equating the two expressions, we get:

$$\begin{aligned}&(1 + g_1 z^{-1} + g_2 z^{-2} + \ldots\ldots + g_{M-1} z^{-(M-1)})(1 + \alpha z^{-1})\\ &+ \mu(g_{M-1} + g_{M-2} z^{-1} + \ldots\ldots + g_1 z^{-(M-2)} + z^{-(M-1)})(1 + \beta z^{-1}) z^{-1}\\ &= 1 + h_1 z^{-1} + \ldots\ldots + h_{M+1} z^{-(M+1)}\end{aligned} \quad (16)$$

Since the corresponding coefficients of the two sides are equal, and $\mu = (1+\alpha)/(1+\beta)$, then the following equations hold:

$$\begin{bmatrix}f_1\\ f_2\\ f_3\\ .\\ .\\ .\\ f_{M+1}\\ f_{M+2}\end{bmatrix} = \left(\underbrace{\begin{bmatrix}\alpha & 0 & 0 & . & . & 0\\ \alpha & 1 & 0 & . & . & 0\\ 0 & \alpha & 1 & . & . & 0\\ . & . & . & . & . & .\\ 0 & 0 & 0 & . & \alpha & 1\\ 0 & 0 & 0 & . & 0 & \alpha\\ 0 & 0 & 0 & . & 0 & 0\end{bmatrix}}_{(M+2)\times M} + \mu\underbrace{\begin{bmatrix}-1-\beta & 0 & . & 0 & 0 & 0\\ 0 & 0 & . & 0 & 0 & 1\\ 0 & 0 & . & 0 & 1 & \beta\\ . & & . & 1 & \beta & 0\\ . & . & . & . & . & .\\ 1 & \beta & . & 0 & 0 & 0\\ \beta & 0 & . & 0 & 0 & 0\end{bmatrix}}_{(M+2)\times M}\right)\begin{bmatrix}1\\ g_1\\ g_2\\ .\\ .\\ .\\ g_{M-1}\end{bmatrix} - \begin{bmatrix}-1\\ h_1\\ h_2\\ h_3\\ .\\ .\\ h_M\\ h_{M+1}\end{bmatrix} = \underbrace{\begin{bmatrix}0\\ 0\\ 0\\ .\\ .\\ .\\ 0\\ 0\end{bmatrix}}_{(M+2)\times 1} \quad (17)$$

where f_i denotes the equation in row i. Newton's method is used to solve for g_1, …, g_{M-1}, α,β and μ from the above equations. Then r_m's can be derived, as below.

Step 1: Initialize $\beta^{(0)}=0.6$ (may be different for some sounds and subjects), then $\mu^{(0)}=h_{M+1}/\beta^{(0)}$, $\alpha^{(0)}=\mu^{(0)}(1+\beta^{(0)})-1$. According to Eq. (17), $g_1^{(0)}, g_2^{(0)}, \ldots\ldots, g_{M-1}^{(0)}$ can be solved from:

$$\left(\underbrace{\begin{bmatrix} 1 & 0 & . & . & 0 & 0 \\ \alpha & 1 & . & . & 0 & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & . & . & \alpha & 1 \\ 0 & 0 & . & . & 0 & \alpha \end{bmatrix}}_{M\times(M-1)} + \mu \underbrace{\begin{bmatrix} 0 & 0 & . & . & 0 & 1 \\ 0 & 0 & . & . & 1 & \beta \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 1 & \beta & . & . & 0 & 0 \\ \beta & 0 & . & . & 0 & 0 \end{bmatrix}}_{M\times(M-1)}\right) \begin{bmatrix} g_1 \\ g_2 \\ . \\ . \\ g_{M-1} \end{bmatrix} = \begin{bmatrix} h_1-\alpha \\ h_2 \\ . \\ . \\ h_{M-1} \\ h_M-\mu \end{bmatrix} \quad (18)$$

Step 2: **C**onstruct the Jacobian matrix:

$$J = \begin{bmatrix} \frac{\partial f_1}{\partial \alpha} & \frac{\partial f_1}{\partial \beta} & \frac{\partial f_1}{\partial \mu} & \frac{\partial f_1}{\partial g_1} & . & . & \frac{\partial f_1}{\partial g_{M-1}} \\ . & . & . & . & . & . & . \\ \frac{\partial f_{M+2}}{\partial \alpha} & \frac{\partial f_{M+2}}{\partial \beta} & \frac{\partial f_{M+2}}{\partial \mu} & \frac{\partial f_{M+2}}{\partial g_1} & . & . & \frac{\partial f_{M+2}}{\partial g_{M-1}} \end{bmatrix} \quad (19)$$

According to Eq. (17), then:

$$J = \begin{bmatrix} 1 & -\mu & -1-\beta & 0 & 0 & 0 & . & . & 0 & 0 \\ 1 & 0 & g_{M-1} & 1 & 0 & 0 & . & . & 0 & \mu \\ g_1 & \mu g_{M-1} & g_{M-2}+\beta g_{M-1} & a & 1 & 0 & . & . & \mu & \beta\mu \\ g_2 & \mu g_{M-2} & g_{M-3}+\beta g_{M-2} & 0 & a & 1 & . & \mu & \beta\mu & 0 \\ . & . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . & . \\ g_{M-2} & \mu g_2 & g_1+\beta g_2 & 0 & \mu & \beta\mu & . & . & a & 1 \\ g_{M-1} & \mu g_1 & 1+\beta g_1 & \mu & \beta\mu & 0 & . & . & 0 & a \\ 0 & \mu & \beta & \beta\mu & 0 & 0 & . & . & 0 & 0 \end{bmatrix} \quad (20)$$

Step 3: Given the values of $\alpha^{(k)}$, $\beta^{(k)}$, $\mu^{(k)}$, and $g_1^{(k)}, \ldots\ldots, g_{M-1}^{(k)}$, their new values are calculated:

$$\begin{bmatrix} \alpha^{(k+1)} \\ \beta^{(k+1)} \\ \mu^{(k+1)} \\ . \\ . \\ g_{M-1}{}^{(k+1)} \end{bmatrix} = \begin{bmatrix} \alpha^{(k)} \\ \beta^{(k)} \\ \mu^{(k)} \\ . \\ . \\ g_{M-1}{}^{(k)} \end{bmatrix} - J(\alpha^{(k)},\ldots,g_{M-1}{}^{(k)})^{-1} \begin{bmatrix} f_1(\alpha^{(k)},\ldots,g_{M-1}{}^{(k)}) \\ f_2(\alpha^{(k)},\ldots,g_{M-1}{}^{(k)}) \\ f_3(\alpha^{(k)},\ldots,g_{M-1}{}^{(k)}) \\ . \\ . \\ f_{M+2}(\alpha^{(k)},\ldots,g_{M-1}{}^{(k)}) \end{bmatrix} \quad (21)$$

Step 4: Repeat step 3 until the increment is small enough. Now, we have the solutions of $g_1, \ldots, g_{M-1}$.

Step 5: Construct polynomial $G_{M-1}(z)=1+g_1z^{-1}+\ldots+ g_{M-1}z^{-(M-1)}$. Then, derive the reflection coefficients $r_{M-1},\ldots,r_1$ of the tube model from the polynomial $G_{M-1}(z)$ using Eq. (12). The relative areas $S_1,\ldots, S_M$ of the VTAF can then be derived using Eq. (5).

During phonation, a glottis may never be 100% closed. Thus, an H_{VTF} estimate may contain the effect of $r_g<1$. Next, we investigate the distortion effects of incomplete glottal closures on VTAF estimates obtained assuming $r_g=1$.

3. EFFECTS OF INCOMPLETE GLOTTAL CLOSURES

The effects of incomplete glottal closures on VTAF estimates obtained using the above method are investigated via computer simulations. Given a VTAF and r_g, $H_{GVTF}(z)$ and $H_{VTF}(z)$ can be constructed using Eqs. (8) and (9). From the constructed $H_{GVTF}(z)$ and $H_{VTF}(z)$, VTAF estimates are derived using the method in section 2.4. The difference between the VTAF obtained from $H_{VTF}(z)$ and that from $H_{GVTF}(z)$ is due to the effect of the incomplete glottal closure. VTAFs measured using MRI [7] are used to construct $H_{GVTF}(z)$ and $H_{VTF}(z)$. The sectional length of given VTAFs is L/M=0.396825 cm. Thus, the observable frequency range of VTFs and GVTFs is $F_s/2=Mc/4L=$ 22.05 kHz. For the VTAF of /ɑ/, the parameters of $r_{lip}(z)$ are calculated to be $\alpha=-0.5225$ and $\beta=0.5939$; for /i/, $\alpha=-0.2824$ and $\beta=0.6468$. The frequency responses of r_{lip}, $r_{lip}(z)$, the constructed $H_{VTF}(z)$ and $H_{GVTF}(z)$ with $r_g=0.99$ and 0.95, and the VTAF estimates for /ɑ/ and /i/ are plotted in Figs. 3 and 4. The VTAFs are normalized relative to the maximal cross-sectional areas. The simulations show that 1) the VTAFs derived from the $H_{VTF}(z)$ using the method in section 2.4 are the same as the original ones (see Figs. 3(c) and 4(c), in which squares and dots are co-centered); 2) for vowels with large lip openings (e.g., /ɑ/, /e/), to obtain reasonable VTAF estimates from $H_{GVTF}(z)$, r_g needs to be greater than 0.95; 3) for vowels with small lip openings (e.g., /i/, /u/) to obtain reasonable VTAF estimates from $H_{GVTF}(z)$, r_g needs to be greater than 0.9995. This means that VTAF estimates of vowels produced with smaller lip openings are more distorted by incomplete glottal closures than those with larger lip openings. This is explained below.

From Fig. 1, H_{GVTF} and H_{VTF} are related as:

$$H_{GVTF}(f)=H_{VTF}(f)/(1+Z_{VT}/Z_g) \quad (22)$$

where Z_{VT} is the impedance looking from the back end of the vocal tract into the vocal tract. From Fig. 1:

$$Z_{VT}=P_1/U_g=H_{VTF}P_1/U_{lip}=Z_{lip}H_{VTF}P_1/P_{lip} \quad (23)$$

Thus, Z_{VT} becomes large at the resonance frequencies of the $H_{VTF.}$ It is known that the H_{VTF} resonates more strongly when the lip opening is smaller because a smaller lip opening has smaller radiation resistance, which has a smaller damping effect in the H_{VTF}. Thus, given the same Z_g, the smaller the lip opening, the greater Z_{VT}/Z_g, and the more different the H_{GVTF} is from the H_{VTF}, and consequently, the more different the VTAF derived from the H_{GVTF} is from the VTAF derived from H_{VTF}.

To obtain more accurate VTAF estimates, both distortion effects caused by $r_g<1$ and r_{lip} should be eliminated. Assuming r_g=constant<1, then the H_{GVTF} denominator in Eq. (8) is an $(M+1)^{th}$ order polynomial in z^{-1} with a leading term of unity. The coefficients of $z^{-1},\ldots, z^{-(M+1)}$ in Eq. (8) can be estimated from speech signals. However, from the M+1 coefficients, one cannot determine the M+2 unknowns: r_g, $r_1,\ldots,r_{M-1}$, α, and β. To determine $r_1,\ldots,r_{M-1}$, it is necessary to know r_g or $r_{lip}(z)$. Obviously, to measure the lip opening area and know $r_{lip}(z)$ is more feasible than to measure the glottal area and know r_g. Eliminating both effects of $r_g<1$ and frequency-dependent r_{lip} on VTAF estimates is future work.

4. RESULTS FROM SPEECH SOUNDS AND CONCLUSION

Our method was applied to vowel sounds produced by several female and male subjects, and yields good VTAF estimates for large-lip-opening vowel sounds, but distorted VTAFs for small-lip-opening sounds, just as simulated above. The results obtained from /ɑ/ produced by a female and a male subject are presented here. The signals sampled at F_s=44.1 kHz are shown in Figs. 5(a) and 6(a). The VTFs were estimated over closed glottal phases (the intervals marked using solid lines in (a) and (b) of Figs. 5 and 6) using the method in [3]. The VTF order is M+1=39 and 46 for the female and the male subjects. The frequency responses of the VTF estimates are shown in (c) of Figs. 5 and 6. The derivatives of the glottal waves obtained by inverse filtering the speech signals using the VTF estimates are shown in (b) of Figs. 5 and 6. The VTAFs derived from the VTF estimates using our method are normalized relative to their maximal values and are

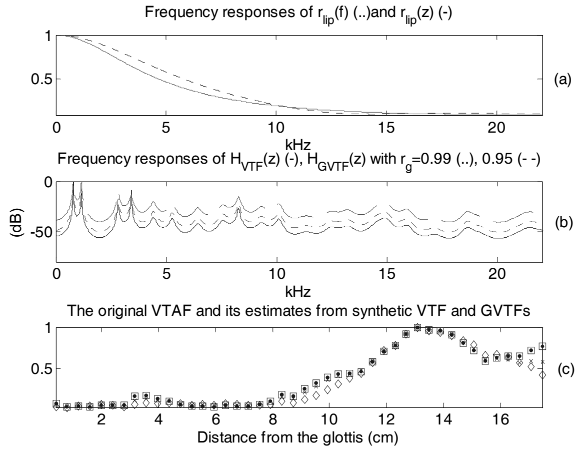

Fig. 3. Vowel /ɑ/: (a) the frequency responses of r_{lip} (broken line) and its IIR model (solid line); (b) the frequency responses of synthetic VTF (solid line) and GVTF with r_g=0.99 (dotted line), 0.95 (broken line); (c) VTAF from MRI (dots), its estimates from VTF (□) and from GVTFs with r_g=0.99 (x), 0.95 (◊).

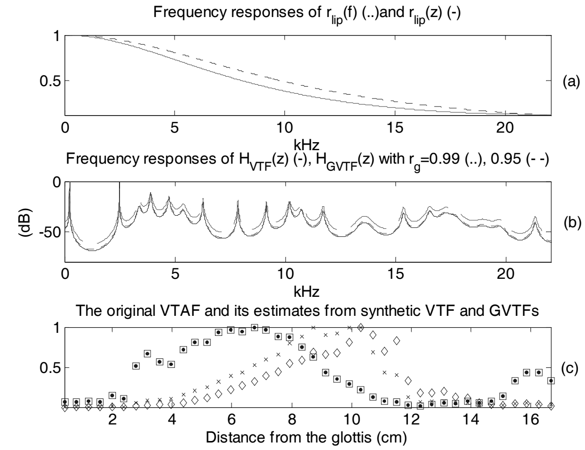

Fig. 4. Vowel /i/: (a) the frequency responses of r_{lip} (broken line) and its IIR model (solid line); (b) the frequency responses of synthetic VTF (solid line) and GVTF with r_g=0.99 (dotted line), 0.95 (broken line); (c) VTAF from MRI (dots), its estimates from VTF (□) and from GVTFs with r_g=0.99 (x), 0.95 (◊).

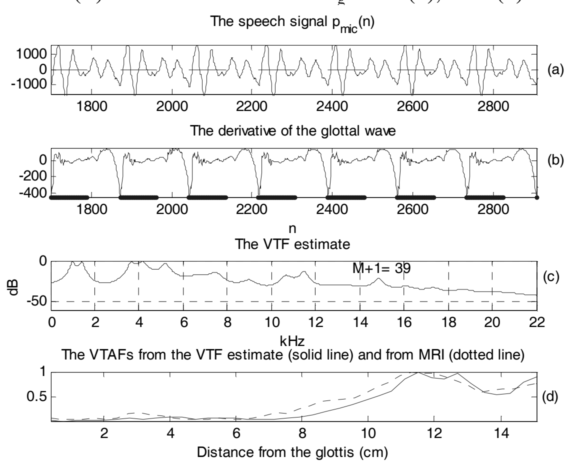

Fig. 5. The speech signal /ɑ/ by a female subject, the estimate of the derivative glottal wave, the VTF estimate, the VTAF estimate (solid line), and the VTAF from MRI (dotted line).

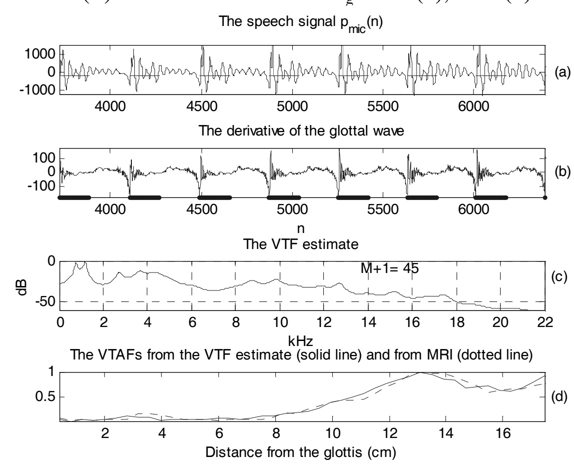

Fig. 6. The speech signal /ɑ/ by a male subject, the estimate of the derivative glottal wave, the VTF estimate, the VTAF estimate (solid line), and the VTAF from MRI (dotted line).

shown using solid lines in (d) of Figs. 5 and 6. They are very similar to the normalized VTAF of /ɑ/ measured for an unknown male subject (the dotted line in plots (c) of Figs. 5 and 6). The agreement between the results from speech signals with those from simulations implies that our VTF models are realistic. The agreement between the VTAFs obtained using our method with that obtained using MRI implies that our methods for obtaining VTF and VTAF estimates from speech signals are correct. For vowel sounds produced with small lip openings, improvements on VTAF estimates can be obtained after both effects of incomplete glottal closures and frequency-dependent lip reflection coefficients in the VTF estimate are eliminated, which requires that the lip-opening areas be known.

5. REFERENCES

[1] B. S. Atal, and L. Hanauer, "Speech Analysis and Synthesis by Linear Prediction of the Speech Wave," *J. Acoust. Soc. Amer.,* Vol. 50, No. 2 (part 2), p 637-655, 1971.

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

[3] H. Deng, R. K. Ward, M. P. Beddoes, and M. Hodgson, "Estimating Vocal-Tract Area Functions from Vowel Sound Signals over Closed Glottal Phases," *IEEE ICASSP,* Montreal, Canada, 2004, Vol. I, pp. 589-593.

[4] T. V. Ananthapadmanabha and G. Fant, "Calculation of True Glottal Flow and Its Components," *Speech Communication* Vol. 1 pp. 167-184, 1982.

[5] Rabiner, L. R., and R. W. Schafer, *Digital Processing of Speech Signals*, Prentice-Hall, New Jersey, 1978.

[6] Kinsler, Lawrence E. *Fundamentals of Acoustics*. John Wiley &Sons, inc., 2000.

[7] B. H. Story, I. R. Titze, and E. A Hoffman, "Vocal Tract Area Functions from Magnetic Resonance Imaging," *J. Acoustic Soc. Am.*, Vol. 100, No. 1, pp. 537-554, 1996.