A vocal-tract model of American English /l/

Zhaoyan Zhang and Carol Y. Espy-Wilson

Citation: The Journal of the Acoustical Society of America **115**, 1274 (2004); doi: 10.1121/1.1645248
View online: http://dx.doi.org/10.1121/1.1645248
View Table of Contents: http://asa.scitation.org/toc/jas/115/3
Published by the Acoustical Society of America

A vocal-tract model of American English /l/

Zhaoyan Zhang[a] and Carol Y. Espy-Wilson[b]
Department of Electrical and Computer Engineering, University of Maryland, College Park, Maryland 20742

(Received 29 May 2003; revised 22 November 2003; accepted 1 December 2003)

The production of the lateral sounds involves airflow paths around the tongue produced by the laterally inward movement of the tongue toward the midsagittal plane. If contact is made with the palate, a closure is formed in the flow path along the midsagittal line. The effects of the lateral channels on the sound spectrum are not clear. In this study, a vocal-tract model with parallel lateral channels and a supralingual cavity was developed. Analysis shows that the lateral channels with dimensions derived from magnetic resonance images of an American English /l/ are able to produce a pole–zero pair in the frequency range of 2–5 kHz. This pole–zero pair, together with an additional pole–zero pair due to the supralingual cavity, results in a low-amplitude and relatively flat spectral shape in the $F3-F5$ frequency region of the /l/ sound spectrum. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1645248]

PACS numbers: 43.70.Bk, 43.70.Fq [AL] Pages: 1274–1280

I. INTRODUCTION

Research on vocal-tract acoustics requires accurate models of sound propagation through complex geometries formed by the articulators. The production of /l/ sounds involves one or two lateral channels which are airflow paths around the tongue produced by laterally inward movement of the tongue toward the midsagittal plane (Stevens, 1998). In cases where the tongue makes contact with the palate, the airflow above the tongue terminates at the contact point along the midsagittal line, giving rise to a ''cul-de-sac'' supralingual side branch to the main airflow pathways of the lateral channels (Fig. 1). When two lateral channels are formed, they split at a point determined by the approach of the tongue dorsum to the teeth, and join again anterior to the lingual–alveolar contact.

The effects of these geometric features on the acoustics of the vocal tract are not clear. In particular, the lateral liquid /l/ is generally characterized by pole–zero clusters around 2–5 kHz in its spectrum. Figure 2 shows a spectrogram of the word ''bell'' produced by a male speaker and Fig. 3 compares spectra during the vowel /ɛ/ and the following /l/. The resonances around 2200 and 3200 Hz during the vowel ($F3$ and $F4$) are absent during the /l/ due to zeros. In fact, the /l/ spectrum is fairly flat between 1600 and 3400 Hz. Although details differ, this scenario holds true for both ''dark'' and ''light'' allophones of /l/ (Lehman and Swartz, 2000). Due to the existence of reported data, this paper will concentrate primarily on accounting for /l/'s as produced in American English. However, the issue of accounting for the acoustics of lateral channels is general to the class of lateral sounds.

There are several possible source(s) of the zeros during /l/ sounds, and their relative contribution to the acoustic spectrum for /l/ remains unclear. One possible source is the midsagittal supralingual side branch (Fant, 1970; Stevens, 1998; Bangayan *et al.*, 1999; Narayanan and Kaun, 1999). An additional source of zeros may be produced when there are two lateral channels and they are of different lengths. Prahler (1998) has shown that uniform lateral channels that have the same length produce a pole–zero pair at the same frequency location so that they cancel each other. Thus, the net result is an all-pole spectrum. If, on the other hand, the lateral channels are of different lengths, then the poles and zeros will be at different frequencies. Specifically, Prahler has shown that when the lateral channels are uniform, but of different lengths, the combined length of the lateral channels needs to be around 16 cm long to produce zeros in the region around 2 kHz. This required length of the lateral channels, however, is much higher than that measured from MRI data (Narayanan *et al.*, 1997; Zhang *et al.*, 2003) or speculated by Fant (1970). As we discuss below, such long lateral channels may not be needed to produce a zero in the $F3-F5$ region if we assume more realistic lateral channels with nonuniform areas.

In this paper, we develop a simple tube vocal-tract model for /l/ that includes both a supralingual side branch and one or two lateral channels. Even though MRI data obtained from a male speaker are used to determine dimensions, the purpose of this study is not to explore the specific vocal-tract shape of a particular individual. Instead, our purpose is to investigate the possible sources of the pole–zero clusters in American English /l/ and the frequency ranges in which they are likely to occur.

II. THEORETICAL BACKGROUND

A. Vocal-tract acoustic response

There have been considerable efforts in the development of computer vocal-tract models. Maeda (1982) developed a time-domain method to calculate the vocal-tract acoustic response function for the vowels and nasal sounds, in which the nasal tract is modeled as a side branch to the vocal tract. This method was modified by Jackson *et al.* (2001) to accommodate an additional side branch needed to model the sublingual cavity of American English /r/'s. To understand better the production of American English /l/, Prahler (1998) developed a vocal-tract model that included uniform lossless

[a]Electronic mail: zhaoyan@glue.umd.edu
[b]Electronic mail: espy@glue.umd.edu

 0001-4966/2004/115(3)/1274/7/$20.00 © 2004 Acoustical Society of America

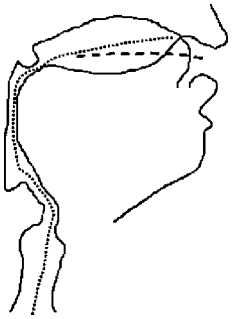

FIG. 1. Schematic of airflow paths above (dotted line) and around the tongue (dashed line) during /l/ sound production. Tracing is from MRI data collected during the sustained /l/ sound produced by the subject used for this study.

lateral channels, but no side branches. Until this study, there was no vocal-tract model that included both side branches and lateral channels, as in the case of /l/ sounds.

In this study, a new frequency-domain model for the vocal-tract acoustic response (VTAR) was developed. The vocal tract is decomposed into various modules such as single tubes, branching, and lateral channels. For a single-tube module the input and output pressures and volume velocities are related by a transfer matrix

$$\begin{bmatrix} p_{\text{in}} \\ U_{\text{in}} \end{bmatrix} = K \begin{bmatrix} p_{\text{out}} \\ U_{\text{out}} \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} p_{\text{out}} \\ U_{\text{out}} \end{bmatrix}, \tag{1}$$

where A, B, C, and D are the coefficients of the transfer matrix K, and depend on the properties of the air and the vocal-tract wall.

The transfer matrix can be calculated using the transmission-line model. The single tube is simulated as a concatenation of cylindrical sections with lengths far less than the acoustic wavelength. Assuming plane-wave propagation, each cylindrical section is represented by an analog circuit as shown in Fig. 4. The transmission-line model has been discussed extensively in many studies (cf. Flanagan, 1972). The exact expressions for each circuit element used in this study are given as follows:

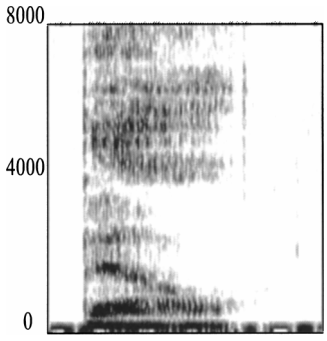

FIG. 2. Spectrogram of word "bell."

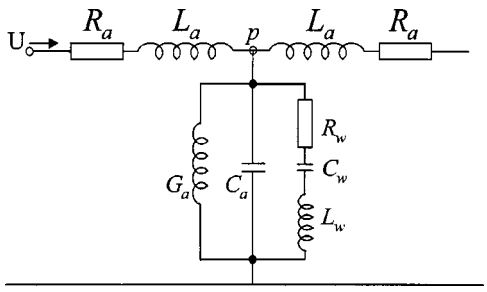

FIG. 4. Transmission line representation of the vocal tract.

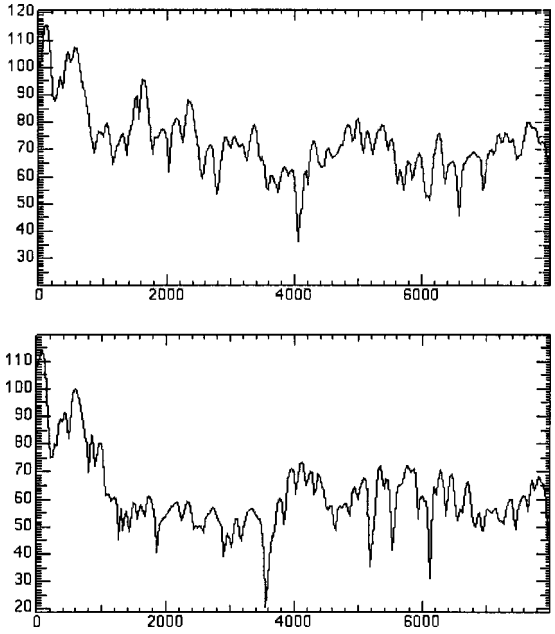

FIG. 3. Spectra during the vowel /ε/ (top figure) and the following /l/ (bottom figure).

$$R_a = \frac{lP}{2S^2}\sqrt{\frac{\omega\rho\mu}{2}}, \quad L_a = \frac{\rho l}{2S}\omega,$$

$$C_a = \frac{lS}{\rho c^2}\omega, \quad G_a = Pl\frac{\eta - 1}{\rho c^2}\sqrt{\frac{\lambda\omega}{2C_p\rho}}, \tag{2}$$

$$L_w = \frac{m}{lP}\omega, \quad R_w = \frac{b}{lP}, \quad C_w = \frac{lP}{k}\omega,$$

where ω is the angular frequency of interest, l is the length of the cylindrical section, S is the tube area, P is the cylindrical circumference, ρ is the air density, c is the speed of sound, μ is the viscosity, λ is the coefficient of heat conduction, η is the adiabatic constant, C_p is the specific heat of air at constant pressure, and m, b, and k are the mass, mechanical resistance, and the stiffness of the wall per unit area of the tube, respectively. This model includes the losses due to the flow viscosity, heat conduction, and vocal-tract wall vibration. The overall transfer matrix of the single tube is simply the product of the transfer matrix of each cylindrical section.

If a branching configuration is present in the vocal tract [Fig. 5(a)], such as the coupling of the nasal tract, a sublingual cavity, or supralingual cavity to the vocal tract, an extra

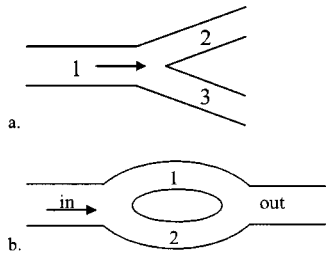

FIG. 5. Models for (a) tube branching and (b) lateral channels.

branch-coupling matrix is used to relate the state variables across the branching point (Sondhi and Schroeter, 1987)

$$\begin{bmatrix} p_1 \\ U_1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1/Z_3 & 1 \end{bmatrix} \begin{bmatrix} p_2 \\ U_2 \end{bmatrix}, \quad \text{or} \quad \begin{bmatrix} p_1 \\ U_1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1/Z_2 & 1 \end{bmatrix} \begin{bmatrix} p_3 \\ U_3 \end{bmatrix}, \tag{3}$$

where Z_2 and Z_3 are the input impedances of the side branches 2 and 3, respectively.

The lateral channels are modeled as two single channels joined together at both ends. Therefore, the two lateral channels have the same input and output pressures [Fig. 5(b)]. Assume for each lateral channel that the input and output state variables are related by

$$\begin{bmatrix} p_{\text{in},i} \\ U_{\text{in},i} \end{bmatrix} = K_i \begin{bmatrix} p_{\text{out},i} \\ U_{\text{out},i} \end{bmatrix}. \tag{4}$$

Applying boundary conditions and flow continuity, simple algebraic manipulation leads to a relationship between the input and output of the lateral channels

$$\begin{bmatrix} p_{\text{in}} \\ U_{\text{in}} \end{bmatrix} = \begin{bmatrix} \dfrac{A_1B_2+A_2B_1}{B_1+B_2} & \dfrac{B_1B_2}{B_1+B_2} \\ C_1+C_2-\dfrac{(D_1-D_2)(A_1-A_2)}{B_1+B_2} & \dfrac{D_1B_2+D_2B_1}{B_1+B_2} \end{bmatrix} \times \begin{bmatrix} p_{\text{out}} \\ U_{\text{out}} \end{bmatrix}. \tag{5}$$

The entire vocal tract is modeled by combining the appropriate modules and multiplying the individual transfer matrices in an order corresponding to their geometric location. This modeling results in a single equation relating the pressures and volume velocities at the glottis and the lips

$$\begin{bmatrix} p_g \\ U_g \end{bmatrix} = K \begin{bmatrix} p_l \\ U_l \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} p_l \\ U_l \end{bmatrix}. \tag{6}$$

The acoustic response function can be calculated as

$$20 \log_{10} |U_l / U_g| = 20 \log_{10} (1/(CZ_l + D)), \tag{7}$$

where Z_l is the radiation impedance at the lips.

B. Code validation

The VTAR model was first validated against Maeda's model (1982) for the simplest case of vowel production. Maeda's model was chosen because it is publicly available and widely accepted. A comparison of the calculated formants for the high front vowel /i/ as in "heed" using Maeda's model and VTAR, with and without radiation load, shows close agreement, within 4 percent. The side branch module was validated against Jackson *et al.*'s (2001) model and complete agreement was obtained.

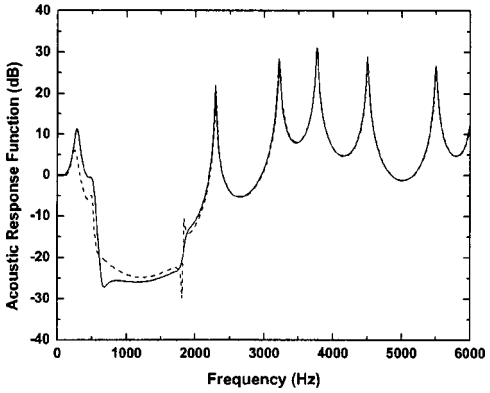

FIG. 6. Comparison between the vocal-tract acoustic response function predicted from Maeda's code (solid line) and VTAR (dashed line), for nasalized vowel /i/, with zero radiation impedance.

Since there are no direct data or models for the lateral channels, data for nasalized vowels were used in this study to validate the lateral channel module. Like /l/'s that are produced with two lateral channels, nasalized vowels are produced with two airflow paths, the oral tract and the nasal tract. In addition, the posterior end of the two paths has the same pressure, assuming plane-wave propagation. The difference between the production of nasalized vowels and /l/ sounds is that, for the nasal sounds, the two paths generally have different radiation impedances and therefore different termination pressures, while the two lateral channels of /l/ have the same termination pressure. Therefore, if the radiation impedance is neglected, nasalized vowels can be modeled using the lateral channel model. Data for a nasalized vowel /i/ and Maeda's simulation code were used in the validation. Figure 6 shows the comparison between the vocal-tract acoustic response functions predicted using Maeda's code and that obtained from the VTAR program. Zero radiation was imposed at both the nose and mouth. The agreement is very good, although there is some mismatch in the spectral amplitude around the first zero frequency (around 600 Hz). These errors are expected in regions around resonances and antiresonances where losses are important. The possible source of error is the different loss formulations used in the time-domain method (Maeda's model) and the frequency domain method VTAR.

C. Uniform lateral channels

For lateral channels consisting of uniform lossless tubes, the input and output pressure and volume velocity of each tube are related by (Kinsler *et al.*, 2000)

$$\begin{bmatrix} p_{\text{in},i} \\ U_{\text{in},i} \end{bmatrix} = \begin{bmatrix} \cos(-kl_i) & -\dfrac{j\rho c}{S_i} \sin(-kl_i) \\ \dfrac{S_i}{j\rho c} \sin(-kl_i) & \cos(-kl_i) \end{bmatrix} \begin{bmatrix} p_{\text{out},i} \\ U_{\text{out},i} \end{bmatrix}, \tag{8}$$

where k is the wave number, and S_i and l_i are the area and length of the ith tube, respectively. Neglecting the radiation at the lateral channels outlet, substitution of Eq. (8) into Eq. (5) yields

$$\frac{U_{\text{out}}}{U_{\text{in}}}=\frac{S_1 \sin kl_2+S_2 \sin kl_1}{S_1 \cos kl_1 \sin kl_2+S_2 \sin kl_1 \cos kl_2}. \tag{9}$$

The terms in the right-hand side of Eq. (9) have specific physical meanings. The two terms in the numerator represent the volume velocity at the outlets of the two channels. Similarly, the two terms in the denominator represent the inlet volume velocity of the two channels. Assuming $S_2>S_1$ allows Eq. (9) to be rewritten as

$$\frac{U_{\text{out}}}{U_{\text{in}}}=\frac{2 \sin k\dfrac{l_1+l_{2,z}}{2}\cos k\dfrac{l_1-l_{2,z}}{2}}{\sin k(l_1+l_{2,p})}, \tag{10}$$

where $l_{2,z}$ and $l_{2,p}$ are the effective zero and pole length of the second lateral channel, respectively, and are defined as,

$$\sin kl_{2,z}=(s_1/s_2)\sin kl_2$$
$$\sin k(l_1+l_{2,p})=\sin kl_1 \cos kl_2+(s_1/s_2)\sin kl_2 \cos kl_1. \tag{11}$$

They represent the effective lengths the second channel needs to be to have volume velocities canceling out with the volume velocity from the first channel at the outlet (therefore a zero) and inlet (therefore a pole), respectively.

Two types of zeros in the acoustic response function of the lateral channels can be identified from Eq. (10). The first set of zeros is at multiples of the frequencies $c/(l_1+l_{2,z})$. This frequency corresponds to the total-length resonance, at which most of the input airflow circulates in between the two channels. The second set of zeros is at odd multiples of the frequencies $c/[2(l_1-l_{2,z})]$. This set of frequencies corresponds to the half-wavelength antiresonance frequency of approximately the difference in the lengths of the lateral channels, in which sound waves at the outlet of the lateral channels are 180° out of phase with each other. The second type of zeros normally occurs at high frequencies since they are related to the length differences. However, in the case of two nonuniform lateral channels, the effective length of $l_{2,z}$ can be very different from l_1, yielding zeros that may appear in the low-frequency range of human speech.

In general, the two effective lengths are functions of both the area ratio and length ratio of the two channels. Consider two special cases in which the two channels have the same lengths and the same areas, respectively. For both cases, the two effective lengths are the same and equal to the length of the wider lateral channels. Equation (10) reduces to

$$\frac{U_{\text{out}}}{U_{\text{in}}}=\begin{cases}1/\cos kl & l=l_1=l_2\\ \dfrac{\cos k\dfrac{l_1-l_2}{2}}{\cos k\dfrac{l_1+l_2}{2}} & S_1=S_2.\end{cases} \tag{12}$$

When the lateral channels have the same length but different areas, the second set of zeros does not exist and the first set of zeros is canceled by poles so that the vocal-tract acoustic response function consists only of poles. In this case, the lateral channels do not contribute any zero to the sound spectra, as shown in Prahler (1998). When the lateral channels have the same areas but different lengths, the first set of zeros is also canceled out by poles. However, the second set of zeros remains, and the lateral channels will contribute zeros to the sound spectra.

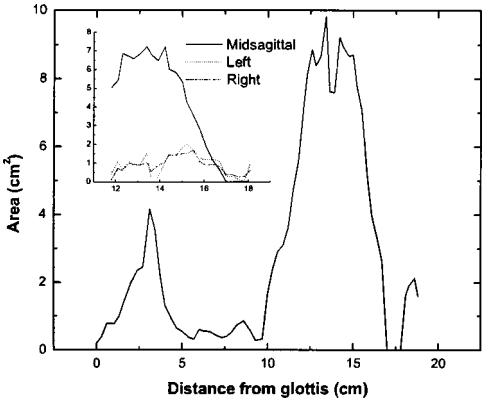

FIG. 7. MRI-derived area functions of the vocal tract, lateral channels (left and right), and supralingual cavity (midsagittal).

It should be noted that lateral channels are usually asymmetric in real speech production. The changes in the acoustic impedance due to the area variation will modify the sound field, and change the phase of the sound waves arriving at the outlets, thereby changing the effective length. The effective lengths of the two channels, due to the area variation, are different even if the two channels have the same lengths [Eq. (11)]. Therefore, the lateral channels always produce pole–zero pairs in real speech production.

III. DATA

Two sets of data were used in the study to illustrate the individual contributions from the supralingual cavity and the lateral channels with reasonable dimensions. The first set of data, which is referred to as the real data or real area functions, is cross-sectional area functions obtained from MRI images of the vocal tract during a sustained /l/ production (Zhang *et al.*, 2003) by the same speaker producing the ''bell'' word shown earlier. The MRI-derived area function is shown in Fig. 7. Notice that the real area function is nonuniform along the length.

The axial length of the midsagittal closure is generally very small, about 1–2 cm long (Narayanan *et al.*, 1997). However, the flow split, therefore the starting point of the lateral channels and the supralingual cavity, may occur at a position posterior to the midsagittal closure. The velocity and the pressure may show different amplitudes and phases, and therefore needs to be modeled acoustically as separate channels. The lateral inward bracing of the tongue may also help the early flow split into two or three channels. In the data processing (Zhang *et al.*, 2003), the vocal-tract cross sections in the region immediately posterior to the midsagittal closure were divided into three regions (the supralingual cavity and

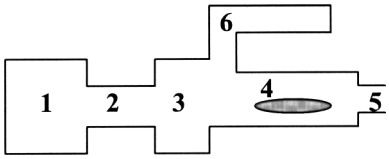

FIG. 8. Simple-tube model of the vocal tract for /l/ sound production.

two lateral channels), and their area functions are also shown in Fig. 7. The supralingual cavity and the lateral channels therefore start posterior at the same location.

The second set of data is the simple-tube data derived from the real MRI-derived area functions. The simple-tube model, as manifested by its name, simplifies the geometry of the vocal tract and its acoustic complexity, and therefore enables us to easily identify the articulatory affiliation of the formants. The simple-tube data were obtained by first dividing the whole vocal tract into several sections. As shown in Fig. 8, the model consists of a back cavity, a pharyngeal constriction, a middle cavity, two lateral channels, a front cavity, and a supralingual cavity modeling the space between the tongue and the palate posterior to the oral closure in the midsagittal line. If the pharyngeal constriction is weak, the first three tubes may be combined into one long back cavity with area perturbations in the middle range (see Stevens, 1998). The area of each section is obtained by averaging the area function over its length. The dimensions of the simple tube model used in this study are shown in Table I. Note that the two lateral channels have the same length in Table I.

Acoustic recordings of a sustained /l/ were also obtained from the same speaker while in supine position at a different time from the MRI recordings. The speaker was attempting to duplicate the sustained /l/ produced during the MRI session.

IV. SIMULATIONS

We consider three simulation cases to help us understand the contributions of the supralingual cavity and the lateral channels in the production of /l/. In the first two simulations, the whole vocal-tract model except the supralingual cavity was used to investigate the possible contributions of the lateral channels alone to the /l/ spectrum. The supralingual cavity was then included in the vocal-tract model in the third simulation and its effects on the spectrum were studied. Simple-tube data were used unless otherwise mentioned.

TABLE I. Dimensions of simple-tube model of /l/ sounds used in simulations. Values in parentheses are used in simulation A for lateral channels of different lengths.

	Length	Area
Back cavity	5.7	1.4734
Pharyngeal constriction	4.3	0.5098
Middle cavity	3.9	5.7064
Lateral channel 1	4.2(5.0)	1.011
Lateral channel 2	4.2(3.4)	0.963
Front cavity	0.7	1.9290
Supralingual cavity	3.1	3.9807

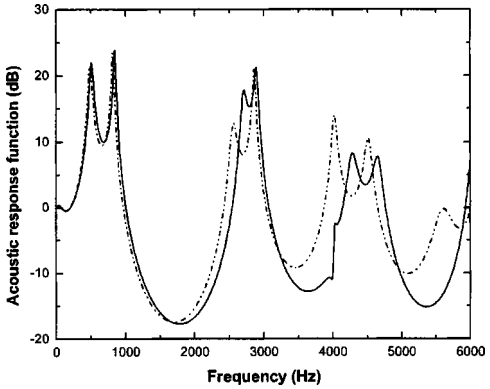

FIG. 9. Vocal-tract acoustic response function for simple-tube model with two uniform lateral channels of different lengths, modeled using lateral channel model (solid line) and with two lateral channels combined into one single tube with an area equal to the combined areas of the two lateral channels (dash-dotted line). Without supralingual cavity.

A. Lateral channels of different lengths

The first simulation considers the case of uniform lateral channels with different lengths (the values in parentheses in Table I). The supralingual cavity was excluded to isolate the effects of the lateral channels on the acoustic response function. Figure 9 shows the calculated vocal-tract acoustic response function. Also shown in the figure is the acoustic response function obtained when the two lateral channels are combined into one single tube with area equal to the sum of the area of the two separate lateral channels. The presence of two lateral channels gives rise to zeros in the acoustic response function, as compared with the all-pole spectrum if the lateral channels are modeled as a single tube. A zero at around 4 kHz is present in Fig. 9, which corresponds to the first type of zero for the total length of 8.4 cm for the lateral channels.

B. Lateral channels of equal length

Figure 10 shows the acoustic response function calculated for the case of uniform lateral channels of equal length. As discussed earlier, uniform channels of equal length do not produce zeros in the spectrum. However, this is not true if the real area functions of the lateral channels are used, as shown in Fig. 10 with the dash-dotted line. Note that a pole–

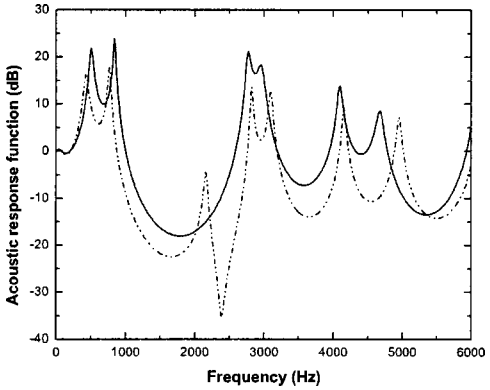

FIG. 10. Vocal-tract acoustic response functions for simple-tube model with two lateral channels of same lengths and uniform (solid line) and nonuniform (dash-dotted line) real area function. Without supralingual cavity.

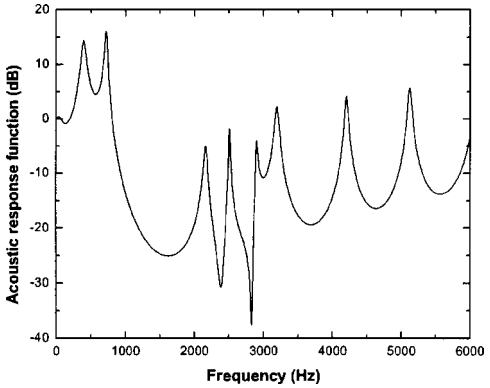

FIG. 11. Vocal-tract acoustic response function for simple-tube model with two nonuniform lateral channels of equal length and a uniform supralingual cavity.

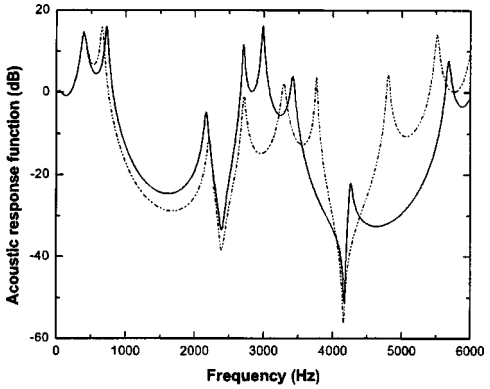

FIG. 12. Vocal-tract acoustic response function for the simple-tube model with two nonuniform real lateral channels and a nonuniform supralingual cavity (solid line), and vocal-tract acoustic response function when the real area function for the entire vocal tract (dash-dotted line) was used.

zero pair appears around 2.4 kHz in the resulting spectrum. The pole–zero pair occurs because the effective lengths of the two channels are unequal due to the nonuniform area functions, according to Eqs. (10) and (11). The nonuniform lateral channels of equal length in this case are essentially the same as uniform lateral channels of different length, as in simulation A. Therefore, zeros are produced even when the two nonuniform lateral channels are of the same length.

Note that the two lateral channels in both simulations A and B have the same combined length (8.4 cm). However, the zero frequency in simulation B (around 2.4 kHz seen in Fig. 10) is much lower than that in simulation A (around 4 kHz seen in Fig. 9) where the lateral channels were also of unequal length but uniform area. The nonuniformity of the area function effectively increases the combined length, resulting in a lower frequency for the zero. This implies that, in real speech production, the lateral channels do not need to be as long as 8 cm to produce zeros in the 2–3-kHz range. It is interesting to note that the zero at 2.4 kHz corresponds roughly to a first type of zeros of a combined length of 14.6 cm if the lateral channels are of uniform area. This length is close to that required in Prahler's work (16 cm) to produce zeros in the range of 2–3 kHz.

C. Supralingual cavity

The addition of the supralingual cavity to the model results in the acoustic response function shown in Fig. 11. Uniform simple-tube data were used for the whole vocal tract except the lateral channels, for which real area data was used. The supralingual cavity, as a side branch, gives rise to an additional zero around 2.8 kHz. The frequency of the zero corresponds to the first quarter-wavelength resonance of a supralingual cavity of length 3.1 cm. The two pole–zero pairs cluster in the range of 2–3 kHz, which coincides with the range of $F3-F4$ for vowels. The prominence of the nearby formants ($F3$ and $F4$) is weakened by the presence of the zeros.

In the production of the lateral sounds, the supralingual cavity is, of course, not uniform in area. In most cases, the supralingual cavity has a tapering area function towards the lingual–alveolar constriction. This tapering actually decreases the effective length of the supralingual cavity, thereby increasing its corresponding zero frequencies, as shown by the solid line in Fig. 12. In this case, the real area function of the supralingual cavity was used.

Also shown in Fig. 12 is the acoustic response function calculated when the real area function was used for the whole vocal tract. The pole values are slightly changed as the real area function reflects more fine features which the simple-tube data fail to include. These changes in the locations of the poles and zeros, although small, significantly change the spectral shape, especially the relative prominence of the pole–zero cluster in the range of 2–4 kHz.

V. DISCUSSION

The simple tube model of American English /l/ presented in this paper shows that both the supralingual cavity and the presence of two lateral channels result in pole–zero clusters around $F3$ and above. In the spectrum of the word-final /l/ in ''bell'' shown in Fig. 3, the pole–zero clusters weakened $F3$ and $F4$, thereby resulting in a fairly flat spectrum between 1600 and 3400 Hz. In Fig. 13, we show the long-term average power spectrum of the acoustic data of the sustained /l/ produced from the same speaker. In the sustained /l/ spectrum, $F3$ and $F5$ have been weakened considerably by zeros and $F4$ is still evident around 3700 Hz. One possible source of the variability in the /l/ spectral shape is

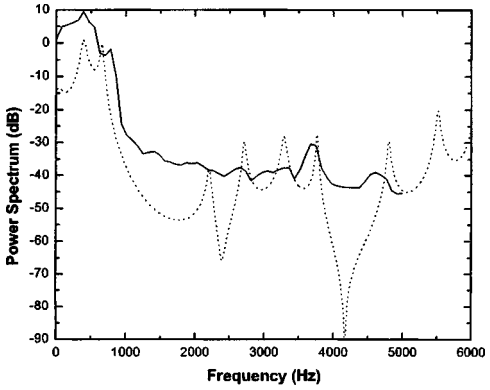

FIG. 13. Measured spectrum of sustained /el/ sound (solid line), and predictions from measured vocal-tract area function with corrections for effects of the source and the radiation (dash-dotted line).

the relative movement of the poles and zeros. Figure 12 has shown that the movement of the poles significantly affects the overall spectral shape. In the simulations, we have shown that the asymmetry of the lateral channels can contribute zeros to the spectrum. The zero frequencies depend on the whole area functions of the lateral channels, and therefore are very sensitive to the changes in either the channel length or the area at one point. This sensitivity will lead to a large variability in the zero location and therefore in the overall /l/ spectral shape.

To compare the simulations with the real power spectrum, the sound spectrum is estimated by adding the effects of the source and radiation as discussed in Stevens (1998) to the acoustic response function calculated by VTAR (dash-dotted line in Fig. 12). The resulting spectrum is shown in Fig. 13 in comparison with the measured power spectrum of the sustained /l/. The two spectra compare well in their spectral shape, especially in the frequency locations of the resonances and the zeros. The first and second formants and the peaks around 3700 and 4600 Hz in the sound power spectrum are accurately predicted by VTAR. The pole–zero cluster in the predicted spectrum that is due to the lateral channels coincide with the ripples in the same range (2000 to 3500 Hz) in the real power spectrum. The predicted zero due to the supralingual cavity occurs during the relatively flat part of the spectrum in the region between 3700 and 4600 Hz. This agreement is not surprising given that the acoustic modeling is based on the same speaker's MRI data. In fact, the agreement between the power spectrum and the model prediction is quite good considering that the acoustic and MRI data were recorded at different times. The main difference is the prominence of resonances and antiresonances above $F2$. The predicted bandwidths of these poles and zeros are much narrower than those observed in the power spectrum. It is possible that some losses related to the air passage through the narrow lateral channels are either not modeled or are underestimated in Eq. (2). In addition, there were some errors involved in estimating the lengths and areas of the lateral channels and supralingual cavity, as discussed in Sec. III. Particularly, since the speaker's teeth have not been superimposed on the MRI slices, the cross-sectional areas of the lateral channels may be overestimated. Smaller areas of the lateral channels will result in more loss, which could reduce further the prominence of the poles and zeros in the cluster region.

VI. CONCLUSION

A vocal-tract model with parallel lateral channels and a supralingual cavity has been developed in this study. Theoretical analysis based on uniform lateral channels shows that lateral channels may produce two types of zeros. The first type of zeros occurs at frequencies at which most of the input airflow circulates around the lateral channel loop. The second type of zeros occurs at frequencies at which the sound waves at the outlet of the two channels are 180° out of phase with each other. Simulations based on both simple-tube data and real data (for the lateral channels only) show that lateral channels of small length (less than 4 cm) can produce zeros in the range of 2–3 kHz, if they have nonuniform areas. The addition of the supralingual cavity introduces another pole–zero pair in the 2–5 kHz range. Thus, this model shows that both the lateral channels and the supralingual cavity are possible causes of the pole–zero clusters in the $F3-F5$ region during /l/ sound production.

ACKNOWLEDGMENTS

This work was supported by NIH Grant 1 R01 DC05250-01 and 1 K02 DC00149-01A1. The authors are grateful to Mark Tiede for the MRI data and for his helpful comments during this project. We also thank Suzanne Boyce for helpful comments on an earlier version of this manuscript and Ken Stevens for a helpful discussion in the beginning of this work.

Bangayan, P., Alwan, A., and Narayanan, S. (**1999**). ''From MRI and acoustic data to articulatory synthesis: A case study of the lateral approximants in American English,'' Proceedings of International Congress of Spoken Language Processes, Philadelphia, PA.

Fant, G. (**1970**). *Acoustic Theory of Speech Production* (Mouton, The Hague, Netherlands).

Flanagan, J. L. (**1972**). *Speech Analysis, Synthesis, and Perception* (Academic, New York).

Jackson, M. T., Espy-Wilson, C., and Boyce, S. (**2001**). ''Verifying a vocal tract model with a closed side-branch,'' J. Acoust. Soc. Am. **109**, 2983–2987.

Kinsler, L. E., Frey, A. R., Coppens, A. B., and Sanders, J. V. (**2000**). *Fundamentals of Acoustics* (Wiley, New York).

Lehman, M. E., and Swartz, B. (**2000**). ''Electropalatographic and spectrographic descriptions of allophonic variants of /l/,'' Percept. Mot. Skills **90**, 47–61.

Maeda, S. (**1982**). ''A digital simulation method of the vocal-tract system,'' Speech Commun. **1**, 199–229.

Narayanan, S., Alwan, A., and Haker, K. (**1997**). ''Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. I. The laterals,'' J. Acoust. Soc. Am. **101**, 1064–1077.

Narayanan, S., and Kaun, A. (**1999**). ''Acoustic modeling of tamil retroflex liquids,'' Proceedings of International Congress of Phonetic Society, San Francisco.

Prahler, A. (**1998**). ''Analysis and Synthesis of the American English Lateral Consonant,'' MIT thesis, Cambridge, Massachusetts.

Sondhi, M. M., and Schroeter, J. (**1987**). ''A hybrid time-frequency domain articulatory speech synthesizer,'' IEEE Trans. Acoust., Speech, Signal Process. **35**, 955–967.

Stevens, K. N. (**1998**). *Acoustic Phonetics* (The MIT Press, Cambridge, MA).

Zhang, Z., Espy-Wilson, C. Y., and Tiede, M. (**2003**). ''Acoustic modeling of American English lateral approximants,'' Proceedings of the Eighth Eurospeech Conference.