AN MRI-BASED ARTICULATORY AND ACOUSTIC STUDY OF LATERAL SOUND IN AMERICAN ENGLISH

Xinhui Zhou[1], Carol Y. Espy-Wilson[1], Mark Tiede[2], Suzanne Boyce[3]

[1]Department of Electrical and Computer Engineering, University of Maryland, College Park, USA
[2]Haskins Laboratories and MIT R.L.E., USA
[3]Department of Communication Sciences and Disorders, University of Cincinnati, USA
[1]{zxinhui@umd.edu, espy@umd.edu}, [2]tiede@haskins.yale.edu, [3]boycese@ucmail.uc.edu

ABSTRACT

The production of the lateral sounds generally involves a linguo-alveolar contact and one or two lateral channels along the parasagittal sides of the tongue. The acoustic effect of these articulatory features is not clearly understood. In this study, we compare two productions of /l/ in American English by one subject, one for a dark /l/ and the other for a light /l/. Three-dimensional vocal tract models derived from the magnetic resonance images were analyzed. It was shown that zeros in the vocal tract acoustic response are produced in the F3-F5 region in both /l/ productions, but the number of zeros and their frequencies are affected by the length of the linguo-alveolar contact and by the presence or absence of lateral linguopalatal contacts. The dark /l/ has one zero below 5 kHz, produced by the cross mode posterior to the linguo-alveolar contact, while the light /l/ has three zeros below 5 kHz, produced by the asymmetrical lateral channels, the supralingual cavity and the cross mode posterior to linguo-alveolar contact.

Index Terms— lateral sound, three dimensional vocal tract, magnetic resonance imaging, and finite element analysis

1. INTRODUCTION

The production of the lateral sound /l/ generally involves a linguo-alveolar contact and one or two lateral channels along the parasagittal sides of the tongue blade [1, 2], as shown in **Fig. 1a**. The acoustic effect of these geometric features is not clearly understood. As shown in **Fig. 1b**, the spectrum of /l/ has relatively weak energy in the F3-F5 region. It has been proposed that this weak energy is due to the pole-zero clusters in the F3-F5 region produced by the lateral channels and/or the supralingual space [1, 3, 4], and the complexity of the /l/ spectrum is caused by the variability of the zero's frequency [1]. However these studies were generally based on an assumed area function vocal tract model and were not based on acoustic analysis of the three dimensional (3-D) vocal tract. Given the articulation complexity, a 3-D acoustic analysis of the vocal tract may provide additional insights on the production of /l/, and it may also give guidance on building an area function vocal tract model of /l/.

English speakers typically use two different varieties of /l/, called "dark" and "light", according to context. This paper presents a comparative study of two productions of /l/ in American English by a male native speaker. One is for a dark /l/, and the other is for a light /l/. First, detailed 3-D geometric reconstruction of the vocal
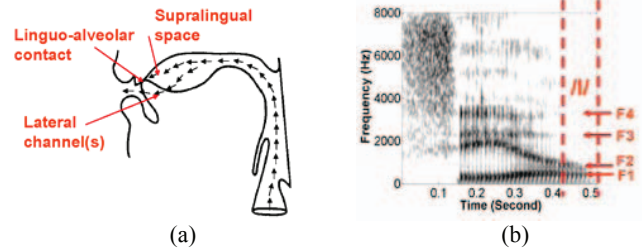

(a)   (b)

Fig. 1. (a) Midsagittal profile of the vocal tract producing /l/, adapted from [1], (b) The spectrogram of the word "feel".

tracts were acquired from magnetic resonance (MR) images. Second, finite element analysis (FEA or FEM) was performed to obtain the acoustic response of the 3-D vocal tract. The wave propagation property at various frequencies was studied to understand the acoustic effects of the lateral channel(s) and the linguo-alveolar contact. Third, area function models were obtained from the 3-D geometry and the resulting acoustic responses were verified against the 3-D acoustic responses. Fourth, simple 3-D vocal tract models were studied to gain additional insights on the acoustic effects of the lateral channels and the linguo-alveolar contact. The focus of this paper is to explain differences between the two /l/'s in number and placement of zeros in the spectrum.

2. MATERIALS AND METHODOLOGIES

2.1. Subject information and data acquisitions

A native American English speaker (subject S2) was selected for this study from our larger study on productions of American English /r/ and /l/ [5, 6]. He produced both a sustained dark /l/ (as in "pole") and a sustained light /l/ (as in "lee") with MR images acquired. The midsagittal MR images of the tongue shapes for these two /l/s are shown in **Fig. 2a**. Note that both of the tongue shapes are just two examples for the /l/ production and there is no strong dividing line between dark /l/ and light /l/ articulatorily [7].

The data collected includes MRI data of the vocal tract for sustained /l/ (sagittal, axial, and coronal slices) and acoustic data recorded in a sound-treated booth for sustained /l/. MR imaging was performed on a 1.5 Tesla G.E. machine. The scanning sequence used was FMPSPGR (Fast MultiPlanar SPoiled GRadient echo) with TR (Time of repetition) 110 ms and TE (Time of echoing) 4.2 ms. The thickness is 3 mm for coronal slices at the linguo-alveolar contact and 5 mm for other slices. The image in-plane resolution is 0.938 mm per pixel.
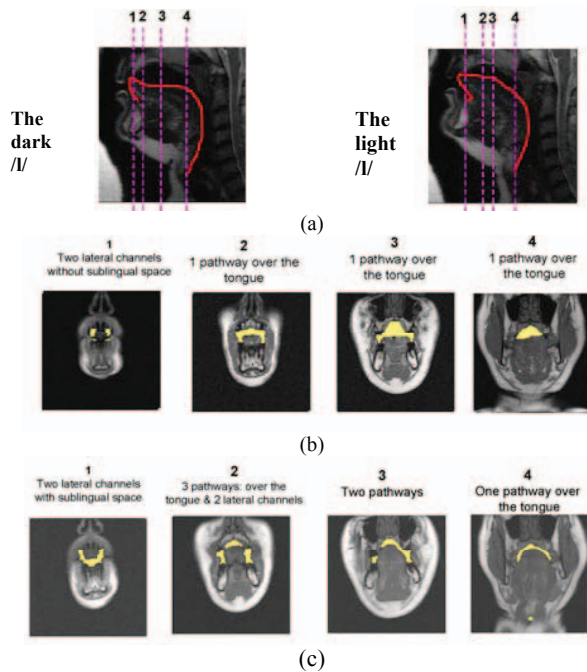
Fig. 2. MR images of the vocal tracts for the dark /l/ and the light /l/. (a) Midsagittal slices (the boundary of the tongue manually drawn in red color for a better visualization of its shape.). (b) Coronal slices of the dark /l/ at different positions specified by dash lines in (a). (c) Coronal slices of the light /l/ (yellow color indicates the airway).

### 2.2. 3-D vocal tract reconstruction and finite element analysis

The medical image processing software MIMICS (Materialise, Inc) was used to process MR images to get a 3-D reconstruction of the vocal tract and the geometry is represented in the STL (STeroLithography) format. The FEM-based harmonic analysis was performed on this geometry using the COMSOL MULTIPHYSICS package, assuming a hard wall and pressure release condition at the lips. The excitation at the glottis was the normal velocity profile of a sinusoidal signal. Details are in [5, 6].

3. RESULTS

3.1. Reconstructed 3-D vocal tract geometries

As shown in **Fig. 2a**, both tongue shapes have a linguo-alveolar contact. However the linguo-alveolar contact for the dark /l/ is established with the tongue tip and that for the light /l/ is established with the tongue blade. The linguo-alveolar contact for the dark /l/ is relatively shorter, as are the lateral channels. Additionally, the tongue dorsum is lowered for the dark /l/, whereas it is raised for the light/l/. Thus, the light /l/ has lateral linguopalatal contacts which make the lateral channels longer and also separate the supralingual space as a side branch.

These differences can be seen in the coronal slices in **Fig. 2b** and **Fig. 2c**. For the dark /l/, there is only one coronal slice (at position 1) having two lateral channels. The other slices have only one air pathway. For the light /l/, the slice at position 1 shows a cross-section of the two lateral channels around the tongue. The slice at position 2 intersects three air pathways which include two

(a) The dark /l/   (b) The light /l/
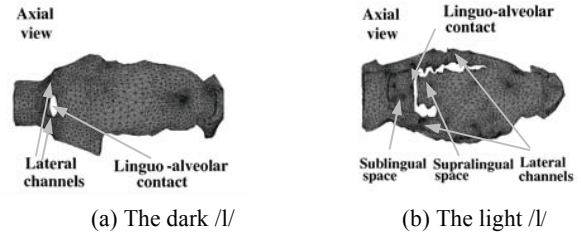
Fig. 3. FEM meshes of the reconstructed 3-D vocal tract. (a) The dark /l/ and (b) The light /l/
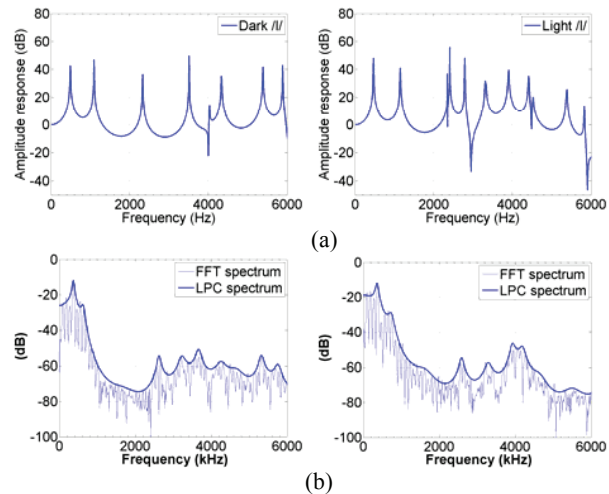
Fig. 4. Left side for the dark /l/ and right side for the light /l/. (a) Acoustic responses based on the 3-D FEM, (b) spectra of sustained /l/ utterance in booth acoustic data

lateral channels and one supralingual space. The slice at position 3 has only one lateral linguopalatal contact on the left, so the supralingual space is connected to one lateral channel on the right side. This means that the light /l/ has asymmetrical lateral linguopalatal contacts and, therefore, two asymmetrical lateral channels with different lengths.

Fig. 3 shows the axial views of the 3-D reconstructed geometries of the vocal tracts for the dark /l/ and the light /l/, respectively. The axial view of the dark /l/ shows a short linguo-alveolar contact which is about 0.8 cm long. The linguo-alveolar contact for the light /l/ is about 1.7 cm long. The axial view of the light /l/ shows two asymmetrical lateral channels (about 4.9 cm long on the right vs. 2.1 cm long on the left) and a separate supralingual space like a side branch.

3.2. FEM-based acoustic responses and wave propagation properties

Based on the reconstructed 3-D geometries, FEM analysis was performed. **Fig. 4** shows the resulting acoustic responses along with the spectra of booth acoustic data. The 3-D acoustic responses of the dark /l/ and the light /l/ have very similar patterns in F1, F2, and F3. However, they have different zeros. The dark /l/ has a zero at 4000 Hz, whereas the light /l/ has zeros at 2350 Hz, 2950 Hz, and 4490 Hz. The zero at 2350 Hz in the light /l/ is hard to detect, because the pole-zero pair are very close to each other. There are some discrepancies in F1-F3 between the 3-D acoustic response and the spectra of the booth acoustic data, which is probably due to

500 Hz   4000 Hz

(a)

500 Hz   2350 Hz

2950 Hz   4490 Hz

(b)
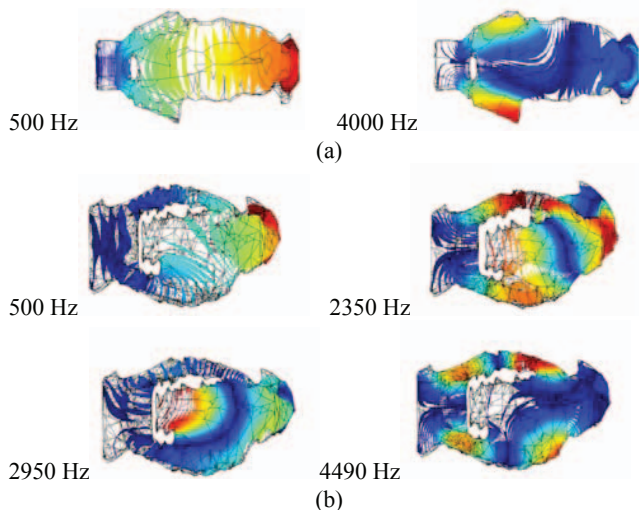
Fig. 5. Pressure isosurfaces plots of wave propagation inside the vocal tracts at different frequencies. (The red color stands for high pressure amplitude, the blue color stands for low amplitude). (a) The dark /l/ and (b) The light /l/.

the subject's inconsistent articulations at different sessions. However, there is a deep valley at about 3980 Hz in the booth acoustic spectrum for the dark /l/, and a deep valley at about 2980 Hz for the light /l/. These deep valleys match well with the corresponding zeros in the 3-D acoustic responses.

In order to understand how the zeros are produced, the wave propagation at different frequencies, indicated by the pressure isosurfaces inside the vocal tract, was studied. The pressure isosurfaces at 500 Hz and 4000 Hz for the dark /l/ are shown in **Fig. 5a**. The wave propagation at 500 Hz is approximately planar. However, at 4000 Hz, a cross mode appears in the region posterior to the contact. The wave at cross-mode propagates towards the two sides of the vocal tract, and hardly comes out from the lips. Therefore, the volume velocity at the lips is extremely small and a zero is produced. **Fig. 5b** shows the pressure isosurfaces for the light /l/ at frequencies 500 Hz, 2350 Hz, 2950 Hz and 4490 Hz (the last 3 are zero frequencies). The asymmetry of the vocal tract caused by the linguo-alveolar contact and the linguopalatal contact make the wave propagation more complex than it is in the case of dark /l/. At 500 Hz, the wave propagation in each branch is approximately planar. At the first zero frequency of 2350 Hz, both branches have approximately planar wave propagation. This zero is attributed to the asymmetry between the two lateral channels. A zero is produced when the volume velocity output of the two lateral channels are almost 180 degrees out of phase. At the second zero frequency of 2950 Hz, the pressure isosurfaces in the supralingual space indicate that the supralingual space functions like a separate side branch. The side branch has almost zero impedance at this frequency and traps all of the energy. Therefore, a zero is produced by the supralingual cavity. At the third zero frequency of 4490 Hz, the cross mode appears just as it does in the dark /l/, which produces the third zero.
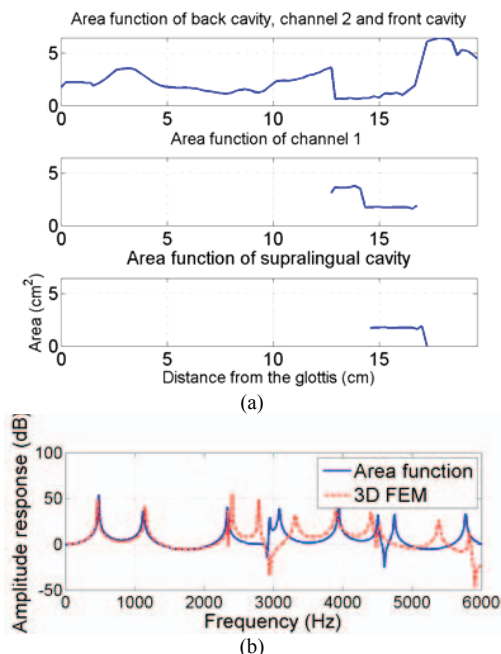
3.3. Area function based vocal tract models

Fig. 6 shows the schematics of the area function vocal tract models used for the dark /l/ and the light /l/, respectively. The supralingual cavity is not in the model for the dark /l/. The area functions have

(a)   (b)
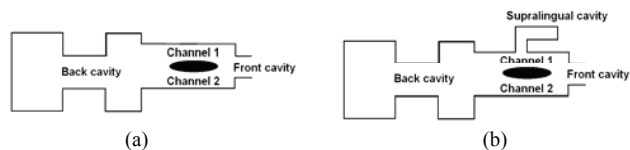
Fig. 6. Schematics of area function vocal tract models for /l/ production of S2 (Each component consists of an area function). (a) The dark /l/, and (b) The light /l/.
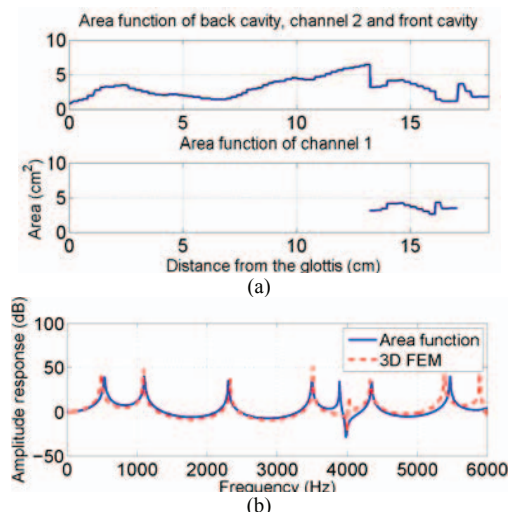
Fig. 7. For the area function vocal tract model of the dark /l/. (a) Area functions, (b) Acoustic responses from area functions and 3-D FEM.

Fig. 8. For the area function vocal tract model of the light /l/. (a) Area functions, (b) Acoustic responses from area functions and 3-D FEM.

been extracted by following the pressure isosurfaces, as described in [6]. However assigning area functions to each component in the model is an empirical process and it should be verified by the

result from the 3-D acoustic analysis. The acoustic response of the area function vocal tract model here was computed by VTAR [3].

For the dark /l/, when the 0.8-cm-long lateral channels are used as the two channel model, the zero at 4000 Hz from 3-D FEM is not reproduced by the area function model. However, when the channels were lengthened to be 3.7 cm long by assigning the cross section area posterior to the contact equally to the two channels, the zero at 4000 Hz is reproduced. This set of area functions and the corresponding acoustic response are shown in **Fig. 7a** and **b**, respectively.

In the area function model for the light /l/, channel 2 is the right lateral channel, and channel 1 consists of the left lateral channel plus part of the supralingual space. There is a supralingual cavity as a side branch of channel 1. The two channels are about 4.5 cm. When the supralingual cavity's length is set as 1 cm, which is intuitive geometrically, there is only one zero at 4000 Hz produced by the area function model. When the supralingual cavity is lengthened to be 3 cm long through sharing the cross-section posterior to the contact with channel 1, two zeros (at 2910 Hz and 4600 Hz) are produced and they are close to the zeros that resulted from the 3-D FEM analysis. This set of area functions and the corresponding acoustic response are shown in **Fig. 8**. It can be proved that the two channels produce the zero at 2910 Hz and the supralingual cavity produces the zero at 4600 Hz.

3.4. Simple 3-D vocal tract models for simulating the lateral channels

To gain more insights into the acoustic effect of the lateral channels in /l/ production, simple 3-D vocal tract models were studied. A simple 3-D vocal tract model is shown in **Fig. 9a**. It is a uniform tube with a rectangular cross section where a 4-cm-long block is positioned in the front to simulate the two lateral channels. It has been shown that a zero does not appear below 6000 Hz when the two channels are symmetrical or only one lateral channel exists. For two asymmetrical channels, the acoustic response has a zero at 4630 Hz for h=4/5H or a zero at 3340 Hz for h=H. This indicates that the two lateral channels with lengths of 4 cm can produce a zero below 6000 Hz, but only when there is a closure or a narrow constriction. A closure can lower the frequency of the zero. **Fig. 9b** shows the acoustic responses at different channel lengths. These simulations are based on the asymmetrical configurations with a closure. When the length varies from 2 cm to 6 cm, the zeros vary from 5130 Hz to 2440 Hz accordingly. The longer channel produces a zero at a lower frequency.

4. SUMMARY

Two productions of sustained American English /l/, one for a dark /l/ and the other for a light /l/, were investigated. Based on the three-dimensional vocal tract models derived from the magnetic resonance images, finite element acoustic analysis was performed. It was shown that both the dark /l/ and the light /l/ have similar patterns in F1-F3. It was also shown that zeros are produced in the F3-F5 region for both /l/ productions, but the number of zeros and their frequencies are affected by the length of the linguo-alveolar contact and by the presence or absence of the lateral linguopalatal contacts. The dark /l/ has one zero below 5 kHz, produced by the cross mode posterior to the linguo-alveolar contact, while the light /l/ has three zeros below 5 kHz, produced by the asymmetrical lateral channels, the supralingual cavity and the cross mode poste-
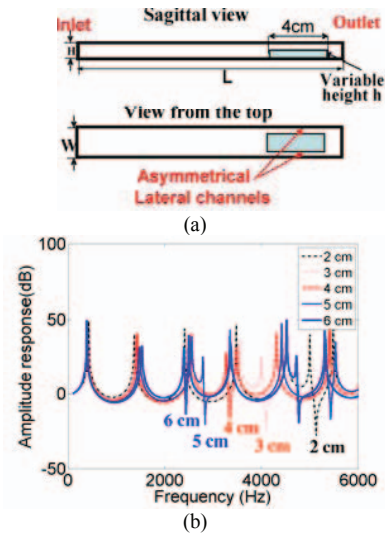

(a)

(b)

Fig. 9. (a) The simple 3-D vocal tract model (H: 1.4 cm, W: 2.8 cm, L: 18 cm, block width: 1.4 cm, block starting location: 5 cm from the outlet, the ratio of lateral channel cross area is 3:1), (b) Acoustic responses of simple vocal tract models with asymmetrical lateral channels at different lateral channels lengths (block height =H).

rior to the linguo-alveolar contact. The results of our simple 3-D vocal tract model analysis show that the lateral channels have to be asymmetrical and 3-6 cm long to get a zero in the region of F3-F5. In addition, a narrow constriction or a closure is also required.

6. REFERENCES

[1] K. N. Stevens, *Acoustic phonetics*. Cambridge, Mass.: MIT Press, 1998.

[2] S. S. Narayanan, A. A. Alwan, and K. Haker, "Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part I. The laterals," *Journal of the Acoustical Society of America*, vol. 101, pp. 1064-1077, 1997.

[3] Z. Y. Zhang and C. Y. Espy-Wilson, "A vocal-tract model of American English /l/," *Journal of the Acoustical Society of America*, vol. 115, pp. 1274-1280, 2004.

[4] A. Prahler, "Analysis and Synthesis of the American English Lateral Consonant," MIT master thesis, Cambridge, Massachusetts., 1998.

[5] X. Zhou, "An MRI-based articulatory and acoustic study of American English liquids sounds /r/ and /l/," PhD dissertation, University of Maryland, College Park, 2009.

[6] X. Zhou, C. Y. Espy-Wilson, M. Tiede, S. Boyce, C. Holland, and A. Choe, "An MRI-based articulatory and acoustic study of 'retroflex' and 'bunched' American English /r/ sounds," *Journal of the Acoustical Society of America*, vol. 123, pp. 4466-4481, 2008.

[7] R. Sproat and O. Fujimura, "Allophonic variation in English /l/ and its implications for phonetic implementation," *Journal of Phonetics*, vol. 21, pp. 291-311, 1993.