Three-dimensional vocal tract modeling of fricatives /s/ and /sh/ for post-glossectomy speakers

Xinhui Zhou*, Jonghye Woo, Maureen Stone and Carol Y. Espy-Wilson

*Corresponding author's address: University of Maryland, College Park, College Park, MD 20740, zxinhui2001@gmail.com

Production of fricatives involves a narrow supraglottal constriction along the vocal tract. Air flows through the constriction, and generates turbulent noise source(s) by impinging on obstacles downstream. In post-glossectomy speakers, production of /s/ and /sh/ is often problematic. It is mainly caused by the tongue surgery which changes tongue properties such as volume, motility, and symmetry, preventing the tongue from creating proper constrictions. The purpose of this study was to gain some insights on how the vocal tracts of abnormal /s/ and /sh/ are shaped and what are their corresponding acoustic consequences. Based on cine magnetic resonance images, we built 3-D vocal tract models for /s/ and /sh/ from two post-glossectomy speakers (one with abnormal /s/ and the other with abnormal /sh/). Due to the missing part of the tongue, the reconstructed vocal tracts are asymmetric with either an air-flow bypass or a side branch formed near the constrictions. Two coupled physics submodels are included in the 3-D FEM acoustic simulation: incompressible potential flow for the mean air flow and aeroacoustics for the distributed noise sources. The resulting acoustic spectra and acoustic roles of air flow bypass or side branch will be discussed. [This study was supported by NIH R01CA133015]

Published by the Acoustical Society of America through the American Institute of Physics
I. INTRODUCTION

Production of fricatives /s/ and /sh/ involves a narrow supraglottal constriction along the vocal tract. When they are produced, air flows through the constriction, and then generates turbulence noise source(s) by impinging on some obstacles downstream (Shadle, 1985; Narayanan et al., 1995; Stevens, 1998). In general, /s/ has a spectral peak at higher frequency than /sh/ and more energy in 3.5-5 kHz (as opposed to 2.5-3.5 kHz) (Jongman et al., 2000; Kent and Read, 2002). In post-glossectomy speakers, the production of /s/ and /sh/ is often problematic (Bressmann et al., 2004). This is mainly caused by the changed tongue properties. The surgery changes tongue volume, motility, and symmetry, preventing the tongue from creating proper constrictions to produce those sounds.

The purpose of this study was to gain some insights on how the vocal tracts of abnormal /s/ and /sh/ are shaped and what are their corresponding acoustic consequences. Based on cine magnetic resonance images, we built 3-D vocal tract models for /s/ and /sh/ from two post-glossectomy speakers (one with abnormal /s/ and the other with abnormal /sh/) and performed acoustic analysis using finite element method.

II. DATABASE AND METHODOLOGIES

A. MR Database

Two post-glossectomy patients (named ‘PT1’ and ‘PT2’ respectively) were used in this study. Both are native American English adult speakers. Both patients had T2 lateral tumors in the lateral margin of the mid section of the tongue (2-4 cm in the largest dimension). They both had primary closures after surgery without radiation or chemotherapy. Their protruded tongues are shown in Fig. 1. A perceptual test discriminating /s/ and /sh/ showed that PT1 had abnormal /s/ and PT2 had abnormal /sh/. The MRI words “a souk” and “a shell” were chosen as the speech task during the MRI session because they take less than 1 second to repeat, which is within the limits of our MRI recording, involve a /s/, which is difficult for glossectomy speakers, and involve little jaw motion maximizing the deformation of the tongue. Cine-MRI data was collected for the MRI words using the following parameters: 3.0 T Siemens Tim Trio; frame rate 26 Hz; in-plane resolution: 1.875 mm/pixel; slice thickness: 6 mm; acquisition of three orthogonal image stacks (sagittal, coronal, and axial). Fig. 2 shows the mid-sagittal MR images of /s/ and /sh/. In both PT1 and PT2, the tongue shapes for /s/ and /sh/ are quite similar.

B. Methodologies

Image processing and vocal tract segmentation

We first created an isotropic volume by applying a super-resolution technique (Woo et al., 2012) on the three orthogonal image stacks (sagittal, coronal, and axial). Then, for teeth compensation, we overlaid the 3-D dental cast shapes extracted from the CT images on top of the MR images. After that, we segmented the airway by thresholding at gray values that were approximately halfway from the air to the tissue near the boundary. Manual correction was also performed at our best guess in regions with over-segmentation or under-segmentation. Based on the segmented vocal tract, we derived the vocal tract surface shapes and saved them in the stereolithography (STL) format for acoustic analysis using finite element method (FEM).

Finite element analysis

We imported the obtained 3-D vocal tract shape in the STL format into the COMSOL MULTIPHYSICS package (Comsol, 2013) and then performed our finite element analysis on it. The mesh for FEM was created using tetrahedral elements as in the STL format. There are two governing equations for the acoustic analysis, one is for incompressible potential flow and the other is for aeroacoustics.

III. RESULTS

A. 3-D vocal tract shape reconstructions of /s/ and /sh/

The 3-D vocal tract reconstructions (/s/ of PT1 and /sh/ of PT2) are shown in Fig. 3. There are two main observations on the difference between patients and normal. First, patients have a larger front cavity due to the reduced tongue volume in the front; Second, the missing tongue tissue may create a bypass of air flow (as in /s/ of PT1) or a side branch (as in /sh/ of PT2). These differences may make the articulatory configuration in patients (such as the constriction location and the front cavity dimension) deviate from its target.
Figure 1. Protruded tongues of two post-glossectomy patients (T2 tumors resected with primary closure, the red lines indicate the midline of the tongues)

Figure 2. Mid-sagittal MR images of /s/ and /sh/ A) /s/ in “a souk”, B) /sh/ in “a

Figure 3. Vocal tract reconstructions of /s/ and /sh/. A) /s/ of PT1, and B) /sh/ of PT2 (From left to right: Midsagittal MR slice, coronal MR slice at the constriction, axial and sagittal view of the 3-D shapes)
B. Acoustic simulation

Numerical acoustic analysis on the extracted 3-D vocal tract shapes are in progress. In our presentation, we will discuss the resulting acoustic responses and the acoustic roles of air flow bypass or side branches formed by the surgery in the post-glossectomy patients.

C. CONCLUSION

3-D vocal tract shapes for /s/ and /sh/ in two post-glossectomy patients were extracted. Due to the missing part of the tongue, the reconstructed vocal tracts are asymmetric with either an air-flow bypass or a side branch formed around constrictions. Two physics submodels are included in the 3-D FEM acoustic simulation: one for incompressible potential flow and the other for aeroacoustics. Our numerical simulation is in progress. The resulting acoustic spectra, its sensitivity to geometry change, and acoustic roles of air flow bypass or side branch will be discussed.

ACKNOWLEDGMENTS

This work was supported by a grant from the National Institutes of Health (grant No. R01CA133015).

REFERENCES