One-dimensional and three-dimensional propagation analyses of acoustic characteristics of Japanese and French vowel /a/ with nasal coupling

Hiroki Matsuzaki1,*, Antoine Serrurier2,†, Pierre Badin2,‡ and Kunitoshi Motoki3,§

1Department of Frontier Information Engineering, Faculty of Advanced Engineering, Hokkaido Institute of Technology, 7–15 Maeda, Teine-ku, Sapporo, 006–8585 Japan
2Department of Parole & Cognition, GIPSA-lab/DPC, ENSIEG, Bat. B, 3ème étage, 961 rue de la Houille Blanche, Domaine Universitaire — BP 46, F-38402 Saint Martin d’Hères cedex, France
3Department of Electronics and Information Engineering, Faculty of Engineering, Hokkai-Gakuen University, 1–1 S-26 W-11, Chuo-ku, Sapporo, 064–0926 Japan

(Received 12 November 2012, Accepted for publication 16 May 2013)

Abstract: In this paper, we describe a comparison of the acoustic characteristics of one-dimensional and three-dimensional models of vocal tracts with nasal coupling. One-dimensional acoustic propagation is computed using an electric analog model. A finite element method is used for three-dimensional acoustic simulation. The comparison of these two approaches involves the vocal-tract shape of two subjects, one Japanese male and one French male pronouncing the vowel /a/ in their native language. Results show that the pole/zero pairs ascribed to the nasal coupling for both simulations appeared at almost the same frequency, at least below 2 kHz. Little difference between the one-dimensional and three-dimensional simulations in the transfer functions for the French subject is observed, since the three-dimensional mesh for the French subject is smoother. An extra pole exists in the transfer function of the three-dimensional model for the Japanese subject, possibly caused by the asymmetric structure of the laryngeal cavity. In the three-dimensional distribution of the active sound intensity vectors for the French subject, sound energy fluxes circulate between oral and nasal cavities coupled in the vicinity of the lips and nostrils.

Keywords: FEM, Japanese and French vowels, Vocal-tract model, 1-D and 3-D wave propagation, Nasal coupling

PACS number: 43.70.Bk [doi:10.1250/ast.35.35]

1. INTRODUCTION

Nasalized vowels have long been investigated in detail using a plane-wave propagation model [1–6]. Moreover, acoustic characteristics of three-dimensional (3-D) nasal tracts [7–9] and vocal tracts with nasal coupling [10,11] have also been studied by a finite element method (FEM).

Although 3-D simulation enables the analysis of acoustic characteristics of vocal tracts with a fine structure, the transfer characteristics below the third formant of the vocal-tract models with a fine structure are almost the same as those of simplified models of the vocal-tract shape [12]. In this frequency region, the one-dimensional (1-D) plane-wave propagation model may be sufficient for analyzing the acoustic characteristics of vocal tracts.

In the context of speech production, this study is a first attempt to assess two different acoustic modeling approaches for nasals. The vocal-tract acoustic models aim at simulating the propagation of an acoustic wave in the vocal tract, from the glottis to the lips and nostrils. Two approaches are assessed in this study: a simplified approach in which plane-wave propagation along the vocal-tract midline is considered, and another in which full 3-D propagation is considered. The comparison of these two approaches involves the vocal-tract shape of two subjects, one Japanese and one French, both pronouncing the vowel /a/ in their native language. This articulation provides the interest of having, in both languages, a significant nasopharyngeal coupling area that allows the study of the effects of nasal coupling on the acoustics of the whole tract.

*e-mail: hiroki@hit.ac.jp
†e-mail: A.Serrurier@soton.ac.uk
‡e-mail: Pierre.Badin@gipsa-lab.inpg.fr
§e-mail: motoki@eli.hokkai-s-u.ac.jp
Note that this nasal coupling is observed for a specific subject [11,13], although ordinary Japanese vowels do not include nasalized vowels. This study also includes an acoustic comparison between the French and Japanese /a/ with nasal coupling. Two 3-D volume meshes of the vocal tract, including oral plus nasal tracts, have been created for the two subjects pronouncing the vowel /a/. Acoustic propagation simulations were then performed for these two meshes.

In the second section of the article, the two methods of obtaining the 3-D meshes of the vocal tract are described. The two acoustic models used in the study are explained in detail in the third section. The simulation results are shown in the fourth section for the oral tract only and for the complete tract with the nasal tract coupled to the oral one. The energy flux distribution is also visualized at a few selected resonance frequencies.

2. VOCAL TRACT MESHING PROCESS

2.1. Japanese Speaker Data

Japanese speaker data are based on three types of MRI data obtained at Brain Activity Imaging Center (BAIC), ATR. The subject is a Japanese male in his thirties. The first set of MRI data is vowel MRI data for obtaining the vocal tract during phonation of the Japanese /a/. The second set is teeth MRI data. The scan parameters were as follows: fast spin echo scan, 25.6 × 25.6 cm field of view, 512 × 512 pixel image size, 9 ms TE, 2,900 ms TR, 90° FA. The slice thickness and the number of slices respectively are 0.25 cm and 26 for the vowel MRI data, and 0.15 cm and 51 for the teeth MRI data. The third set is nasal cavity data. The scan parameters were as follows: RFFAST, 12.8 × 12.8 cm field of view, 512 × 512 pixel image size, 5 ms TE, 12 ms TR, 20° FA, coronal slice plane, 60 slices and 0.2 cm slice thickness.

Three surface meshes (STL format) were constructed using the vowel, nasal cavity and teeth MRI data. Then the surface meshes of the nasal cavity and teeth MRI data were superimposed on the surface mesh of the vowel MRI data. After some modification by manual operation, complete closed surface meshes were constructed.

2.2. French Speaker Data

The 3-D mesh of the French speaker vocal tract was extracted from MRI data in the framework of 3-D articulatory modeling [14]. The subject is a male French native speaker, 43 years old at the time of data recording.

Among a larger corpus, one stack of sagittal MR images was recorded from the subject sustaining the vowel /a/ for about 35 s. A set of 25 sagittal images with a size of 25 × 25 cm, 0.36 cm slice thickness and an interslice distance of 0.4 cm was obtained. The image size is 256 × 256 pixels, giving a resolution close to 0.1 cm/pixel. To improve the tracing, especially in the lateral regions, the 3-D block of voxels formed by the original images was resliced to create 26 new images. These images, which were uniformly spaced from the glottis to the lips, were oriented perpendicular to the main oral tract direction, i.e., axial at the glottis level and turning regularly until becoming coronal at the lip level. The 3-D shapes of the jaw and the hard palate, including the teeth, were extracted from a stack of computer tomography (CT) images, in which, contrary to MRI, the bony structures were clearly visible. These 3-D meshes were resliced according to the MR image planes. The contours were superimposed on the MR images. The contours of the oral tract were then manually traced on the 25 original sagittal images and the 26 perpendicular resliced images, augmented with the bony structure contours, to finally form a discrete 3-D description of the tract. The full 3-D reconstruction process is explained in detail in [14] in the framework of 3-D articulatory modeling. Owing to the extremely low resolution of the block of voxels formed by the MR images the 3-D mesh of the nasal passages, which were rather intricate and had a complex shape, was extracted from the CT images through a similar process. The reconstructed complete oral and nasal tract shapes are shown in Fig. 1.

The framework of 3-D articulatory modeling in which this 3-D meshing process has been done resulted in smooth vocal-tract contours without angular borders. Note that this mesh is smoother than that of the Japanese subject.

3. ACOUSTIC MODELS

3.1. Plane-Wave Acoustic Propagation

In a first approximation, an acoustic model of plane-wave propagation was considered. In this model, transverse propagation modes are neglected up to 5 kHz, and therefore, the propagation is planar and 1-D along the midline of the vocal tract. The area function, i.e., the variation of the area of the tract along this line, is then calculated from both 3-D meshes. The surface meshes are sliced according to a
impedance for the 1-D model of the vocal tract, the radiation. Since the frequency range of interest is limited in the infinite baffle is assumed as a physical model of the quadrupoles where they are supposed to be connected. The nasal tract is represented by a series of quadrupoles in parallel to the oral tract being a series of quadrupoles. The nasal tract are indicated in cyan and green, respectively. Note that the sinuses maxillaries are not modeled for the subjects are shown in Fig. 2. Left and right passages of the lengths of 0.5 cm. The area functions obtained for the two nasal tracts. The piriform sinuses are divided equally into various lengths with a minimum of 0.4 cm and a maximum of 0.8 cm for the nasopharyngeal tracts and also divided into various lengths with a minimum of 0.3 cm and a maximum of 0.6 cm for the Japanese subject. The nasopharyngeal and nasal tracts are divided into various lengths with a minimum of 0.2 cm and a maximum of 0.6 cm for the Japanese subject. The main oral tracts are divided into various lengths with a minimum of 0.2 cm and a maximum of 0.6 cm for the main oral tract only are also shown in the lower right in Figs. 3 and 4. A volume of radiation [16] with a radius of 4 cm, which is spherical in shape, was attached to the face covering the lips and nostrils. A specific acoustic impedance of spherical waves was used as a boundary condition on the round surface of the volume of radiation. The spherical surface is shown in green in the lower right of Fig. 3. The glottis, as the driving surface, was driven with a sine wave. A rigid wall condition was assumed. Sound pressure and particle velocity were computed from the velocity wave. A rigid wall condition was assumed. Sound pressure and particle velocity were computed from the velocity wave.

Finite element meshes are constructed through remeshing of the closed surface meshes using a second-order triangular finite element and fully automatic volume mesh generation using a second-order tetrahedral finite element. The finite element meshes of the 3-D vocal-tract models are shown in Fig. 3 for the Japanese subject and in Fig. 4 for the French subject. The meshes for the main oral tract only are also shown in the lower right in Figs. 3 and 4. A volume of radiation [16] with a radius of 4 cm, which is spherical in shape, was attached to the face covering the lips and nostrils. A specific acoustic impedance of spherical waves was used as a boundary condition on the round surface of the volume of radiation. The spherical surface is shown in green in the lower right of Fig. 3. The glottis, as the driving surface, was driven with a sine wave. A rigid wall condition was assumed. Sound pressure and particle velocity were computed from the velocity potential.

The vocal-tract transfer functions and active sound intensity distributions are obtained using the sound pressure and particle velocity. The transfer function \(H_{3D,VN}\) for the 3-D propagation is defined as

\[
H_{3D,VN} = \frac{U_L + U_N}{U_G} \tag{3}
\]

As a first attempt at comparison between the two acoustic models, all the losses are neglected in this study.

### 3.2. 3-D Propagation

A 3-D FEM was applied to the wave equation in a steady state to obtain the velocity potential. The applied FEM was used to simulate acoustic wave propagation in 3-D vocal-tract models.

The radiation impedance of a circular piston set in an infinite baffle is assumed as a physical model of the radiation. Since the frequency range of interest is limited for the 1-D model of the vocal tract, the radiation impedance \(Z_r\) is approximated in a simple form as

\[
Z_r = \frac{\rho c}{\pi a^2} \left\{ \frac{(ka)^2}{2} + j \frac{8ka}{3\pi} \right\}, \tag{1}
\]

where \(\rho\) is the density of air, \(c\) is the velocity of sound, \(a\) is the radius of the mouth or nostril opening and \(k\) is the wave number [15].

The oral transfer function \(H_{1D,V}\) is computed as the ratio of acoustic flows at the lips \(U_L\) to acoustic flows at the glottis \(U_G\).

\[
H_{1D,V} = \frac{U_L}{U_G} \tag{2}
\]

The complete oral plus nasal transfer function \(H_{1D,VN}\) is computed as the ratio of the sum of the acoustic flows at the lips and at the nostrils \(U_N\) to the acoustic flows at the glottis.

\[
H_{1D,VN} = \frac{U_L + U_N}{U_G} \tag{3}
\]
as a quantity that is proportional to the sound pressure at a distant point.

The simulation was carried out in a driving frequency range of 10 Hz to 5 kHz at intervals of 10 Hz and then 1 Hz in the vicinity of poles and zeros.

4. RESULTS

4.1. Oral Tract

The vocal-tract transfer functions computed with the two models for the oral tract alone are shown in Fig. 5. The transfer function $H_{3D}$ becomes negative infinity under direct current, since $H_{3D}$ has first-order differential characteristics and is proportional to frequency in the low-frequency region. Although $H_{1D}$ and $H_{3D}$ are defined in Eqs. (2)–(4), the frequencies of poles and zeros can be correctly evaluated since the difference between $H_{1D}$ and $H_{3D}$ appears only in the envelope of the transfer characteristics. In the case of the French subject, the two transfer functions are very similar in appearance. On the other hand, in the case of the Japanese subject, large differences between the 1-D and 3-D simulations are observed. In the transfer function of the 3-D simulation, there is, in particular, an extra pole (1,058 Hz) between the first and second poles of the transfer function of the 1-D simulation. Some extra poles or pole/zero pairs above 3 kHz are observed.

4.2. Oral and Nasal Tracts

The vocal-tract transfer functions computed using the
two models for the complete vocal-tract model are shown in Fig. 6. Extra poles or pole/zero pairs related to the nasal tract appear in all of the transfer functions. In the transfer function of the 3-D simulation for the French subject, zeros do not appear.

The first three formant (F1–F3), extra pole and zero frequencies are shown in Table 1. Frequency differences between the 1-D and 3-D simulations are small for the French subject and large for the Japanese subject.

### Table 1  Formant, extra pole and zero frequencies in Hz.

<table>
<thead>
<tr>
<th></th>
<th>Main tract only</th>
<th>With nasal tract</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Japanese subject</strong></td>
<td>1-D</td>
<td>3-D</td>
</tr>
<tr>
<td>P</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Z</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>F1</td>
<td>600</td>
<td>555</td>
</tr>
<tr>
<td>P</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Z</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>F1</td>
<td>1,058</td>
<td>1,067</td>
</tr>
<tr>
<td>Z</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

### 4.3. Sound Energy Flux

Figures 7–9 show 3-D distributions of active sound intensity vectors for French subject. Vectors are magnified 3 times for 1,255 Hz, 0.1 times for 909 Hz and 500 times for 1,581 Hz. The magnitude of the vectors are indicated by the lengths of the arrows and the colors. The vectors are magnified by a suitable scale (indicated by the number following “×” in the figures) for better visualization. Red thick arrows represent gross directions of the active sound intensity vectors. At 689 Hz (F1), the magnitude of the vectors in the oral cavity is almost the same as that in the nasal cavity. The sound energy flux from the glottis bifurcates into the oral and nasal cavities. At 1,255 Hz (F2), the magnitude of
the vectors in the nasal cavity is lower than that in the oral cavity. As shown in Fig. 8, the sound energy flux from the glottis meets the energy flux from the nasal cavity at the top of the pharynx, and starts to flow into the oral cavity. At 909 Hz, the magnitude of the vectors in the nasal cavity is higher than that in the oral cavity. At 1,581 Hz, where a small abrupt increase in the amplitude in Fig. 6 is observed, the magnitude of the vectors in the nasal cavity is slightly higher than that in the oral cavity. The sound energy flux radiates from the left nostril, while the sound energy flux drains into the right nostril. Figure 9 shows the top view of the 3-D distribution of active sound intensity vectors for 1,581 Hz. The magnitude of the vectors in the right nasal tract is lower than that in the left nasal tract. The sound energy flux in the right nasal tract flows from the right nostril to the pharynx and drains into the left nasal tract. This pattern and the corresponding small increase in the transfer function at this frequency might be caused by the slight asymmetry of the nasal passages.

Figure 10 shows 3-D distributions of active sound intensity vectors of the complete vocal-tract model for the Japanese subject. At both 567 Hz (F1) and 1,333 Hz (F2), the sound energy flux in the nasal cavity flows from the nostril to the pharynx, and the magnitude of the vectors in the nasal cavity is lower than that in the oral cavity. These distribution patterns are the same as those at 1,255 Hz for the French subject (see Fig. 7). At pole frequencies of 332 Hz and 840 Hz related to nasal coupling, the sound energy flux from the glottis bifurcates into the oral and nasal cavities, and the magnitude of the vectors in the nasal cavity is higher than that in the oral cavity. These distribution patterns are also the same as those at 909 Hz for the French subject (see Fig. 7).

5. DISCUSSION

In Fig. 5 for the oral tract alone, the formants observed for the French subject are consistent with the formants reported in the literature for /a/. On the contrary, additional peaks are observed for the Japanese subject. The peaks at approximately 700 Hz, 1,200 Hz and 2,600 Hz are the typical formants of /a/. The 3-D simulations show an extra pole at approximately 1,000 Hz. Note that the 3-D mesh of the Japanese subject is more complicated than that of the French subject and the 1-D simulations may miss some resonances owing to the simplification process. The formant at approximately 2,400 Hz is due to the Helmholtz resonance of the lowest cavity of the oral tract, near the glottis (see Fig. 2 with area functions). The pole/zero pair at approximately 3,300 Hz can be ascribed to an interdental space [18]. The formant at approximately 4,400 Hz is probably an oral formant shifted by the presence of a pole/zero pair at approximately 4,700 Hz owing to the piriform sinuses, which are not modeled for the French subject.

In Fig. 6, there are no zeros in the transfer function of the 3-D simulation for the French subject. Indeed, the zeros depend on the position of the sphere considered for the calculation of the transfer function; in this case, the intensities are accumulated on the surface of the whole sphere and the zeros are averaged [17]. In the transfer functions computed for the French subject, the formants are observed at almost the same frequency as in Fig. 5.
The presence of the nasal tract parallel to the oral cavity introduces some pole/zero pairs in the transfer function. These pairs correspond roughly to the resonances of the nasal cavities (approximately 900 Hz, 2,500 Hz and 3,900 Hz). The slight asymmetry of the two nasal cavities may be the origin of the small pole/zero pair at approximately 1,500 Hz (see Sect. 4.3) and continues to have an effect up to the resonances at approximately 3,900 Hz. Similar remarks are true for the Japanese subject’s transfer functions. A first pole/zero pair at approximately 300 Hz is observed, which can be ascribed to the Helmholtz resonances of the sinuses maxillaries (not modeled for the French subject).

The extra pole at 1,058 Hz in the transfer function for the Japanese subject is possibly caused by the extremely curved shape of the laryngeal cavity in the lateral direction since the extra pole does not appear when the curved shape in the larynx is replaced by a straight one. The extremely asymmetrical shape in the larynx may produce an additional pole that may not be represented by only the vocal-tract area function. This matter should be further examined as a topic of future work.

In the 3-D distribution of active sound intensity vectors for the model with nasal coupling, the magnitude of the vectors in the nasal cavity was higher than that in the oral cavity. It seems that this kind of distribution pattern may cause a new pole related to the nasal cavity.

6. CONCLUSION AND PERSPECTIVES

The pole/zero pairs ascribed to the nasal coupling, as well as the formants for both the 1-D and 3-D simulations, appeared at almost the same frequencies, at least below 2 kHz.

Little difference between the 1-D and 3-D simulations in the transfer functions for the French subject was observed. It seems that the slight differences are caused by different radiation boundary conditions of the vocal-tract transfer functions. The 3-D mesh for the French subject is smoother, i.e., the mesh has fewer short branches, and more symmetrical with respect to the lateral direction than those for the Japanese subject. As these differences may be caused by the complexity of the shape, it may be concluded that the 3-D fine structure of the vocal tract is an important factor that characterizes the voice sound, and the smoothing of the vocal-tract shape should be avoided in the 3-D simulation.

In the 3-D distribution of the active sound intensity vectors at F2 for the French subject, a flow of the sound energy flux from the nostrils to the top of the pharynx is also observed, which is similar to the results for the Japanese subject [11]. Experimental measurement is needed to verify the simulation results.

ACKNOWLEDGMENTS

Part of this work has been supported by a research project of the High-Tech Research Center, Hokkaido-Gakuen University and Grant-in-Aid for Scientific Research for (B) No. 18300069 Japan Society for the Promotion of Science.

REFERENCES