3D and plane wave acoustic propagation comparison for a Japanese and a French vowel /a/ with nasal coupling

Hiroki Matsuzaki, Antoine Serrurier, Pierre Badin, Kunitoshi Motoki

3D and plane wave acoustic propagation comparison for a Japanese and a French vowel /a/ with nasal coupling *

○ Hiroki Matsuzaki (Hokkai-Gakuen University), △ Antoine Serrurier, △ Pierre Badin (GIPSA-lab, ICP) and Kunitoshi Motoki (Hokkai-Gakuen University)

1 Introduction

In the context of speech production, this study provides a first attempt to assess two different acoustic modelling approaches. The vocal-tract acoustic models aim at simulating the propagation of the acoustic wave in the vocal tract, from the glottis to the lips and the nostrils. Two approaches are assessed in this study: one simplified approach considering plane wave propagation along the vocal-tract midline, and another one considering full three-dimensional (3D) propagation. 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 presents the interest of having in both languages a significant nasopharyngeal coupling area that allows the study the effects of nasal coupling on the acoustics of the whole tract. Note that this nasal coupling is observed for the Japanese /a/[1, 2] although Japanese does not have nasal vowels. This study gives moreover the occasion to have a first acoustic comparison between French and Japanese /a/ with nasal coupling. Two 3D volume meshes of the vocal tract, including oral plus nasal tracts, have thus been obtained for the two subjects for the vowel /a/. Acoustic propagation simulations were then performed for these two meshes.

In the second section of the article, we describe the two methods to obtain the 3D meshes of the vocal tract. The two acoustic models used in the study are detailed in the third section. The simulation results are described in the fourth section: for the oral tract only and for the complete tract with nasal tract coupled to the oral one. The energy flux is also displayed 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 MRI data is vowel MRI data to obtain the vocal tract during phonation of the Japanese /a/. The second one 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, 2900 ms TR, 90° FA. The slice thickness and the number of slices are 0.25 cm and 26 for the vowel MRI data, and 0.15 cm and 51 for the teeth MRI data. The third one is nasal cavity data. The scan parameters were as follows: RF-FAST, 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.

We constructed three surface meshes (STL format) using the vowel, nasal cavity and teeth MRI data, respectively. And then we superimposed the surface meshes of the nasal cavity and teeth MRI data on the surface mesh of the vowel MRI data. After some process of modifying, we constructed complete closed surface meshes.

2.2 French speaker data

The 3D mesh of the French speaker vocal tract has been extracted from MRI data in the framework of 3D articulatory modelling[3]. The subject is a male French native speaker, about 1.65 m tall and 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 artificially the vowel /a/ during about 35 sec. A set of 25 sagittal images with a size of 25 × 25 cm, 0.36 cm thick, and an inter-slice distance of 0.4 cm was obtained. The image resolution is close to 0.1 cm/pixel, the image size being 256 × 256 pixels. Due to the necessity of accurate contours of the various soft organs of the vocal tract for organ-based articulatory modelling, the contours of the soft tissues were edited manually image by image. This process is rather accurate in the sagittal planes close to the midsagittal one, but less accurate in the lateral regions where the vocal-tract surface becomes tangent to the sagittal planes of the images. To improve the tracing, especially in these lateral regions, the 3D block of voxels formed by the original images has been resliced to create 26 new images. These images uniformly spaced from the glottis to the lips are oriented perpendicular to the main oral tract direction, i.e. axial at the glottis level and turning regularly until coronal at the lip level. One of the main problems of MRI remains however in the impossibility to detect the bony structures, which are particularly important in the buccal cavity with the teeth. This problem has been handled by extracting manually the 3D shape of the jaw and of the hard palate, including teeth, from a stack of computer tomography (CT) images, on which, contrary to MRI, the bony structures are clearly visible. These 3D meshes have been resliced according to the MR image planes: the contours obtained have then been superimposed on the MR images. The contours of the oral tract have then been manually traced on the 25 original sagittal images and the 26 perpendicular resliced images, augmented of the bony structure contours, to form finally a discrete 3D description of the tract. These contours have been used as input of a Surface Reconstruction Software (http://cgal.inria.fr/Reconstruction/submit.htm) to obtain the 3D surface mesh of the oral tract visible on Fig. 1. The full 3D reconstruction process is detailed in [4] in the framework of 3D articulatory modelling. Due to the too low resolution of the block of voxels of the MRI, the 3D mesh of the nasal passages, of rather intricate and complex shape, have finally been extracted from the CT images through a similar process. The complete oral and nasal tract shapes are visible on Fig. 1.

The framework of 3D articulatory modelling in which has been done this 3D meshing process imposed to avoid unrealistic vocal-tract contours like angular borders. Note thus that in comparison to the mesh of the Japanese subject, this mesh can be considered as smoother.

---

* 鼻腔結合を伴う日本語とフランス語母音/a/における 3 次元および平面波音響伝搬の比較．松崎 博季 (北海学園大・工)，セルリエ アントアン，バダン ピエール (GIPSA-lab, ICP)，元木 邦俊 (北海学園大・工)

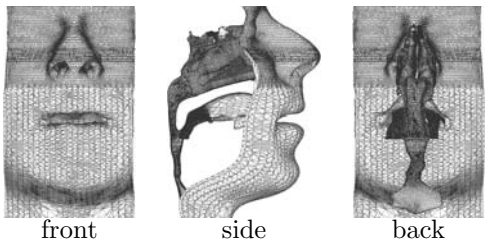

Fig. 1 3D surface mesh for French subject.

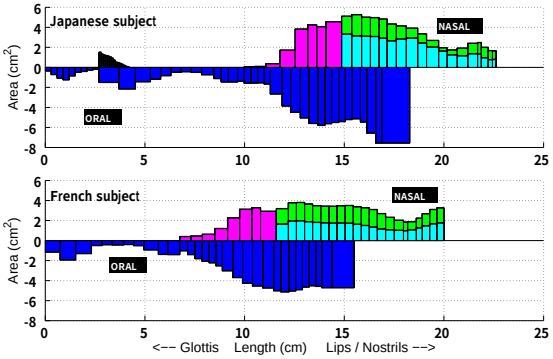

Fig. 2 Area functions

3 Acoustic models

3.1 Plane wave acoustic propagation

In a first approximation, an acoustic model of plane wave propagation was considered. We assume in this model that transverse propagation modes can be neglected until 5 kHz and therefore that the propagation is plane and one-dimensional (1D) 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 for the both 3D meshes. The surface meshes are sliced according to a semipolar grid whose regularly spaced planes are perpendicular to the direction of the vocal tract. The area and the gravity centre of the vocal tract are determined on each slice. The area of each tube of the area function is computed as the average of two consecutive slice areas. The length of the tubes is defined as the distance between consecutive gravity centres. The area functions obtained for the two subjects are visible on Fig. 2. Based on the analogy between acoustic and electric propagation, the acoustic model used for this study is an electric analog, largely used and documented in the literature[5, 6]. Each tube is modelled by an electric quadripole, the oral tract being then a series of quadripoles. The nasal tract is represented by a series of quadripoles in parallel to the buccal cavity quandripoles; the various sinuses are again represented by series of quadripoles in parallel to the tract quadripoles where they are supposed to be connected. The oral transfer function is computed as the proportion between the acoustic flows at the lips and the acoustic flows at the glottis; the complete oral plus nasal transfer function is computed as the proportion between the sum of the acoustic flows at the lips and at the nostrils and the acoustic flows at the glottis. As a first attempt to compare the two acoustic models, all the losses are neglected in this study.

3.2 3D propagation

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

We constructed finite element meshes through a process of re-meshing of the closed surface meshes using a second order triangle finite element and a process of full automatic volume mesh generation using a second order tetrahedral finite element. The finite element meshes of the 3D vocal-tract models are shown in Fig. 3 for the Japanese subject and in Fig. 4 for the French subject. The figures on the left are the meshes for the complete vocal tract with nasal tract (yellow colored part). The figures on the right are the meshes for the main oral tract only. A volume of radiation[7] 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 round surface is only shown in Fig. 3 right as green colored part. 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 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.

We obtained vocal-tract transfer functions and active sound intensity using the sound pressure and particle velocity. The transfer function $H(\omega)$ for the 3D propagation is defined as $H(\omega) = K\left|\sqrt{W_{rad}}/u_g\right|$, where W_{rad} is a radiation power equivalent to the total active intensities on the surface of the volume of radiation, u_g is the source volume velocity, and K is a constant for $H(\omega)$ to be dimensionless[8].

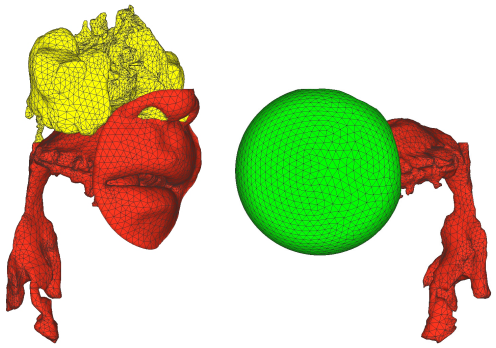
Fig. 3 FE meshes for Japanese subject. Left: complete vocal-tract model with nasal tract, right: model with main oral tract only. Green colored part is round surface of volume of radiation.

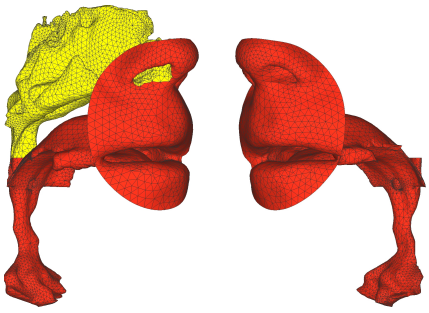
Fig. 4 FE meshes for French subject. Left: complete vocal-tract model with nasal tract, right: model with main oral tract only.

4 Results

4.1 Oral tract

The vocal-tract transfer functions computed with the two models for the oral tract alone are shown in

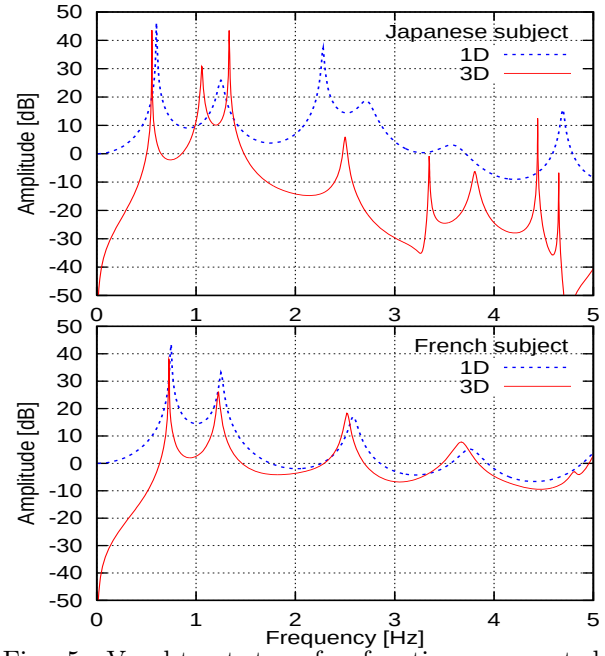

Fig. 5 Vocal-tract transfer functions computed with two models for the oral tract alone.

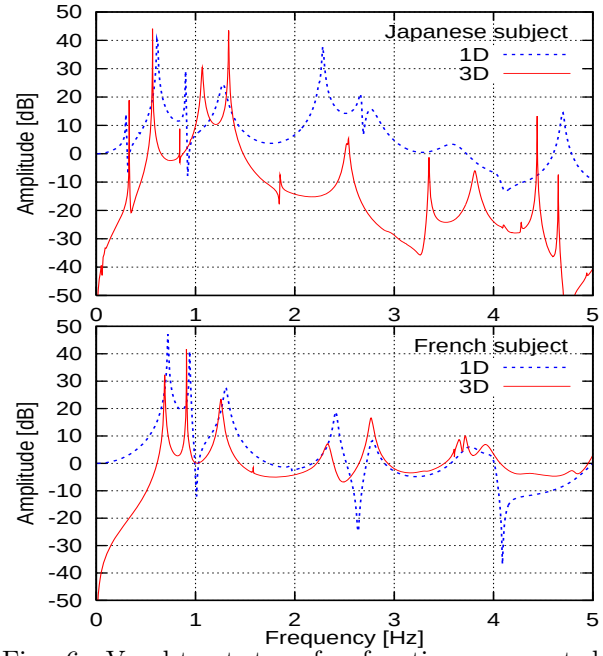

Fig. 6 Vocal-tract transfer functions computed with two models for complete vocal-tract model with nasal tract.

Fig. 5. In the case of the French subject, both the transfer functions resemble closely in appearance. On the other hand, in the case of the Japanese subject, we can see large differences between the 1D and 3D simulations. In the transfer function of the 3D simulation, there is in particular an extra pole (1058 Hz) between the first and second pole of the transfer function of the 1D simulation. We also see some extra poles or pole / zero pairs above 3 kHz. The formants observed on the French subject are coherent with the classical formants reported in the literature for a /a/. On the contrary, we observe more formants for the Japanese subjects. The formants around 700 Hz, 1200 Hz and 2600 Hz are the classical formants of a /a/. The 3D simulations reveal an extra pole around 1000 Hz that is still to be interpreted. Note moreover that the 3D mesh of the Japanese subject is more complicated than the French mesh and the 1D simulations may miss some resonances due to the simplification process. The formant around 2400 Hz is due to the Helmholtz resonance of the lowest cavity of the oral tract, near to the glottis (see Fig. 2 with area functions). Finally, the last formant is probably a shifted oral formant that appears because of the presence of a zero around 5000 Hz due to the sinuses piriformis, that are not modelled for the French subject.

4.2 Oral and nasal tracts

The vocal-tract transfer functions computed with the two models for the complete vocal-tract model are shown in Fig. 6. Extra poles or pole / zero pairs relating to the nasal tract appear in all of the transfer functions. In the transfer function of the 3D simulation for the French subject, we cannot see any zeros. Indeed, the zeros depend on the point of the sphere considered for the calculation of the transfer function: in our case, the intensities are accumulated on the whole sphere and the zeros are averaged. The differences between the 1D and 3D simulations are smaller for the French subject and are larger for the Japanese subject. About the transfer functions computed on the French subject, we observe logically the same formants as the Fig. 5. The presence of the nasal tract in parallel to the buccal cavity introduces some pole/zero pairs in the transfer function. These pairs correspond roughly to the own resonances of the nasal cavities (around 900 Hz, 2500 Hz and 3900 Hz). Interestingly, the light asymmetry of the two nasal cavities is at the origin of the small pole / zero pair around 1500 Hz (see section 4.3) and participates to the resonances around 3900 Hz. Similar remarks can be done about the Japanese subject transfer functions. We observe moreover a first pole / zero pair around 300 Hz that can be ascribed to the Helmholtz resonances of the sinuses maxillaries (not modelled on the French subject).

4.3 Sound energy flux

Fig. 7, 8 and 9 shows 3D distributions of active sound intensity (sound energy flux density) vectors of the complete vocal-tract model for the French subject. The magnitude of the vectors are indicated with a length of the arrows and colors. The vectors are magnified a suitable number (indicated with symbol "×" in the figures) for better visualization. Red colored 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 with that in the nasal cavity. And a sound energy flux from the glottis bifurcates into the oral and nasal cavities. At 1255 Hz (F2), the magnitude of the vectors in the nasal cavity is lower than that in the oral cavity. And 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. It seems that this kind of distribution pattern causes a new pole related with the nasal cavity.

At 1581 Hz, where a small abrupt increase of the amplitude 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

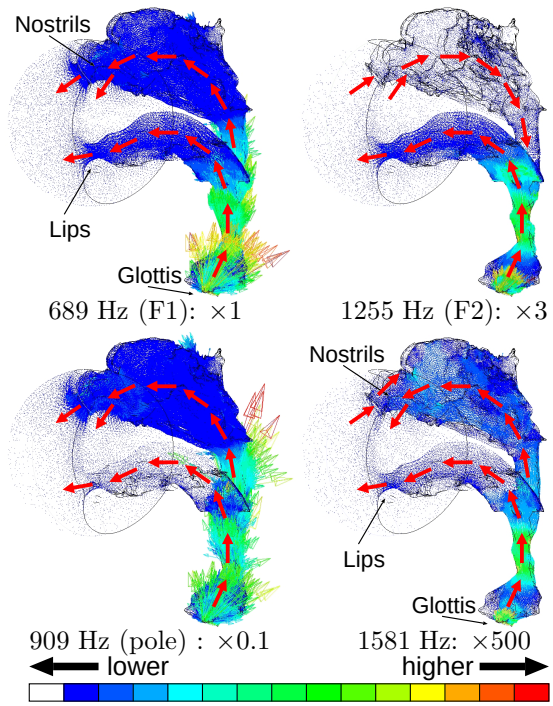

Fig. 7 3D distributions of active sound intensity vectors for French subject. Vectors are magnified 3 times for 1255 Hz, 0.1 times for 909 Hz and 500 times for 1581 Hz.

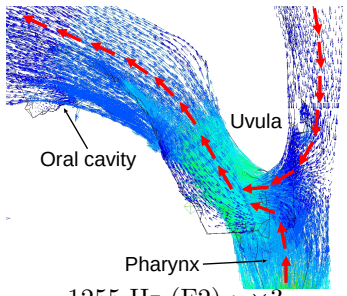

Fig. 8 Elements on large scale near pharynx of distributions of active sound intensity vectors for 1255 Hz.

the right nostril. Fig. 9 shows top view of the 3D distribution of active sound intensity vectors for 1581 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 at this frequency in the transfer function seems clearly be caused by the light asymmetry of the nasal passages.

5 Conclusion & perspectives

The pole / zero pairs ascribed to the nasal coupling for both the 1D and 3D simulations appeared in almost the same frequency at least below 2 kHz.

We could not see large difference between the 1D and 3D simulations in the transfer functions for the French subject. The 3D meshes for the French subject are smoother, i.e. the meshes have few short branches, and more symmetrical laterally than those for Japanese subject. If these differences are caused by the complexity of the shape, it may be said that the 3D fine structure of the vocal tract is an important factor which characterizes the voice sound.

The extra pole at 1058 Hz in the transfer function for the Japanese subject is possibly caused by a laterally curved shape of the laryngeal cavity since this asymmetric structure is not taken into account in the 1D model. The examination of this matter is needed as a future work.

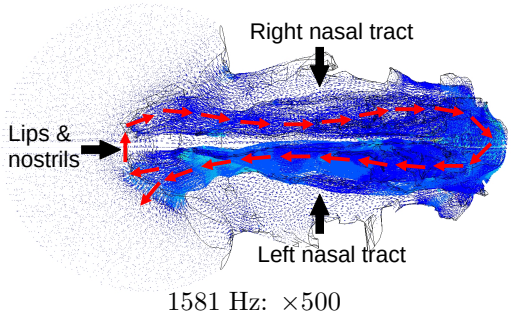

Fig. 9 Top view of 3D distribution of active sound intensity vectors for 1581 Hz. Vectors are magnified 500 times.

In the 3D distribution of the active sound intensity vectors at F2 for the French subject, we could also see the same kind of distributions of the sound energy flux, which flows from the nostrils to the top of the pharynx, with those of a Japanese subject[1].

It is not verified that the results obtained our simulations happen in reality since the sharp peaks and zeros are not observed under the soft wall condition[1]. Experimental measurement will be needed in order to verify the simulation results.

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

Reference

[1] H. Matsuzaki and K. Motoki, “Study of acoustic characteristics of vocal tract with nasal cavity during phonation of Japanese /a/,” Acoust. Sci. Tech., 28(2), 124–127, 2007.

[2] H. Matsuzaki and K. Motoki, “Construction of vocal-tract geometrical model with nasal cavity using MRI data and its acoustic analysis,” Proc. Spring Meet. Acoust. Soc. Jpn., 3-8-17, 227–228, 2007.

[3] A. Serrurier and P. Badin, “A Three-Dimensional Linear Articulatory Model of Velum Based on MRI Data,” Proc. Interspeech 2005, 2161–2164, 2005.

[4] A. Serrurier and P. Badin “Towards a 3D articulatory model of velum based on MRI and CT images,” ZAS Papers in Linguistics, 40, 195–211, 2005.

[5] G. Fant, Acoustic Theory of Speech Production, Mouton, The Hague, 1960.

[6] P. Badin and G. Fant, “Notes on vocal tract computation,” Speech Transmission Laboratory - Quarterly Progress and Status Report, Stockholm, Sweden, 25(2-3), 053–108, 1984.

[7] H. Matsuzaki, N. Miki and Y. Ogawa, “FEM analysis of sound wave propagation in the vocal tract with 3-D radiational model,” J. Acoust. Soc. Jpn. (E), 17(3), 163–166, 1996.

[8] K. Motoki and H. Matsuzaki, “Evaluation of the transfer characteristics of 3-dimensional vocal-tract models,” Proc. Autumn Meet. Acoust. Soc. Jpn., 1-P-16, 2007.