

Time-domain FEM simulation of Japanese and 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

The acoustic analysis of three-dimensional (3-D) geometrical vocal-tract models with nasal coupling for Japanese and French vowel /a/ has been performed by a finite element method (FEM) in frequency domain [1–4]. These previous studies showed that the nasal coupling produces additional peaks below 3 kHz and sound energy circulates through the oral and nasal cavities.

In this paper, we synthesize vowel /a/ of the 3-D vocal-tract models with the nasal coupling for two Japanese and one French subjects using the FEM in time domain. We examine whether there are any differences between the synthesized vowels of the models with and without the nasal coupling by using an ABX discrimination test.

2 Vocal-tract geometrical model

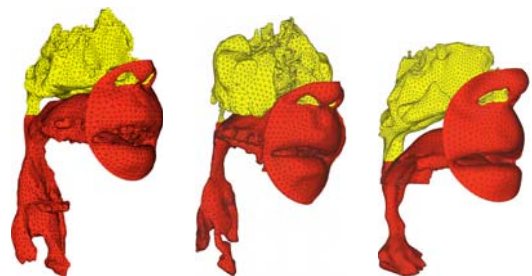
To construct the 3-D vocal-tract geometrical models, we used vowel MRI data of the vocal tract with the nasal coupling during phonation of the vowel /a/ for three subjects A, B and C. The first one is for a Japanese male subject (A) and provided from ATR [1, 2]. This subject has a history of an operation on his paranasal sinuses. The second one is also for a Japanese male subject (B) [3, 4]. The third one is for a French male subject (C) [4] who pronounced the nasalized vowel /a/. We constructed surface meshes using Mimics (Materialise) and then constructed finite element (FE) meshes using HyperWorks (Altair Engineering, Inc.).

Fig. 1 shows FE meshes of the 3-D vocal-tract geometrical models for three subjects. Red colored parts are the main vocal tract. Yellow colored parts are the nasal cavity coupled to the oral cavity with coupling area of 56, 8 and 41 mm² for the subjects A, B and C, respectively. A volume of radiation [5] with a radius of 4 cm, which is spherical in shape, was attached to the face covering the lips and nostrils. The nasal cavity is also coupled to the oral cavity through a space between the lips and the nostrils in a 3-D volume of radiation. We also constructed the FE meshes without the nasal cavity by removing the yellow colored part indicated in Fig. 1.

3 FEM simulation

3.1 Simulation conditions

We used the 3-D FEM applied to the wave equation to synthesize vowel /a/. A rigid wall condition was assumed on the vocal-tract wall. The characteristic acoustic impedance of air was assumed on the round surface of the volume of radiation. As a sound source, we used a glottal waveform estimated from a real voice during phonation vowel /a/ [6]. The subject of the sound source is a Japanese male different from the subjects of the vowel MRI data. Long time average spectra and spectrogram of the sound source are shown in Fig. 2. An upper limit of an effective frequency range of the sound source is about 3 kHz because of a limitation of the estimation method [6].



Subject A Subject B Subject C

Fig. 1 FE meshes for three subjects. Left, center : Japanese subjects, right: French subject. Yellow colored part is nasal cavity.

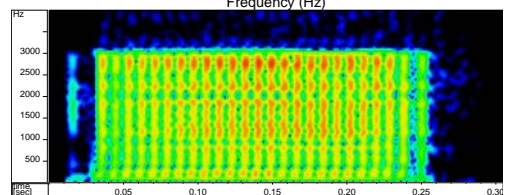
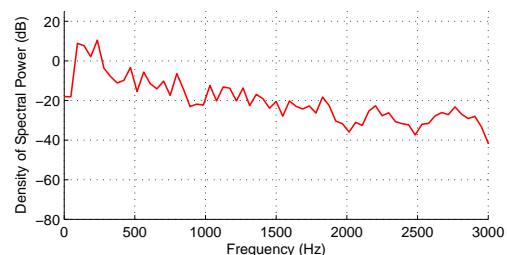


Fig. 2 Long time average spectra (top) and spectrogram (bottom) of sound source estimated from real voice.

Using this sound source, we carried out the FEM simulation for every FE meshes.

3.2 Simulation results

From the results of the FEM simulation, we obtained sound pressure at one node on the round surface of the volume of radiation, in front of the mouth. We translated the data of the sound pressure into audio files with Microsoft WAV file format, mono, 48 kHz sampling, 16 bit quantization and PCM encoding. Long time average spectra over the steady state part of the vowel and spectrograms of the audio files are shown in Fig. 3. The spectra for models with and without nasal cavity are indicated as blue and red colored lines, respectively. Formants F1 to F4 are indicated on the spectrograms as red, green, blue and yellow colored lines, respectively.

Zeros appear in the spectrum near around 750 Hz and 1 kHz for the model with the nasal coupling of the subject A. Compared with the formant frequencies for the model without the nasal coupling, we can observe a split of the F3 formant into two formants F3 and F4, located on both sides of the original oral F3 formant. It is also noted that relatively high amplitude is observed in the F3 and F4 region, and these spectral components do not decrease immediately after the end of the source glottal wave. These characteristics may be understood from the

* 鼻腔結合を伴う日本語とフランス語母音/a/の時間領域有限要素法による音声生成シミュレーション. 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)

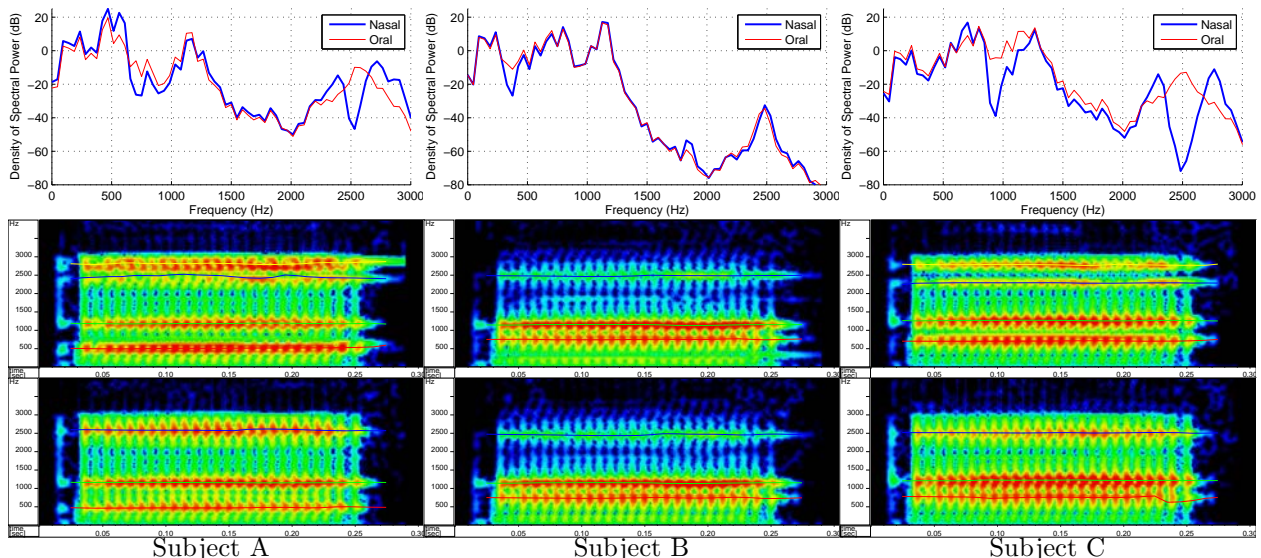


Fig. 3 Long time average spectra (top) and spectrograms (middle for model with nasal coupling and bottom for model without nasal coupling) of synthesized vowel /a/. Left to right: subjects A, B and C.

Table 1 Results of ABX discrimination test.

	#1	#2	#3	#4	#5	#6	#7	#8	#9
A		○	○		○	○	○	○	○
B									○
C	○				○			○	

fact that the present simulation takes into account the sound radiation but assumes no loss factors inside the vocal tract, which makes the acoustic field in the vocal tract highly reverberant. Similar results can be seen for the subject C, but the amplitude decrease after the end of the source glottal wave is rapid. We can also see zeros in the spectrum near around 1 kHz and 2.5 kHz for the model with the nasal coupling for the subject C. There are little differences between the spectra and the spectrograms for the model with and without the nasal coupling for the subject B as there is a smaller coupling area to the nasal tract.

4 ABX discrimination test

To examine whether auditory difference exists between the synthesized vowels of the models with and without the nasal coupling, we performed an ABX discrimination test. Listeners are nine Japanese adult males in their twenties. The listeners were presented the synthesized vowels *A*, next *B* and then *X* with intervals of 0.5 sec. And then they identify *X* as being either *A* or *B* for 5 sec pause. An ABX session consists of 16 trials. A headphone (Sony MRD-2900HD) was used for the test. The test was carried out for each subject’s model.

4.1 Results and discussion

We computed values of a chi-square from the results of the test and then obtained probabilities for each listener. Statistical significant differences at the 5 % level are shown in Table 1 as symbol “○”. There is only 1 significant difference for the subject B, which corresponds well to the fact that there is very small cross-sectional area between the oral and the nasal cavities, and is little difference between the spectra of the model with and without the nasal coupling. 7 and 3 significant differences are obtained for the models of the subject A and C, respectively, possibly caused by the spectral difference around 0.5-1 kHz and 2.5 kHz. Listeners #2, #7 and #9 reported that they perceived longer decaying tone at the end

of the sound for the model with the nasal coupling of the subject A and they judged the difference by the sound. Listener #5, who could perfectly discriminate the differences for the subject A and C, reported an impression of the sound for the model with the nasal coupling being muffled.

5 Conclusion

According to the diverging trends observed on the ABX discrimination test results between the subjects A and B, it is plausible to affirm that the difference on the coupling condition between oral and nasal cavities can be perceived from the signals obtained by the 3-D vocal-tract models with the loss-less condition. Indeed, the results of the ABX discrimination test showed clearly more significant differences for the model constructed from Japanese subject A than for the model constructed from the Japanese subject B. This may be resulting from the fact that the extracted coupling area is clearly larger for the Japanese subject A. Another possible reason is that only Japanese listeners, who are not familiar with nasalized vowels, participated in the test. The same test by French listeners should be also performed. Further investigations need to be led to assess the different results observed between subjects A and C.

Acknowledgments This study has been carried out using “ATR vowel MRI data”. Part of this work has been supported by a research project of the High-Tech Research Center, Hokkai-Gakuen University, Research grant from Hokkai-Gakuen University and Grant-in-Aid for Scientific Research for (B) No. 18300069 Japan Society for the Promotion of Science.

Reference

- [1] Matsuzaki *et al.*, Tech. Rep. IEICE, SP2005-47, 7–12 (2005).
- [2] Matsuzaki *et al.*, Acoust. Sci. Tech., 28(2), 124–127, 2007.
- [3] Matsuzaki *et al.*, Proc. Spring Meet. Acoust. Soc. Jpn., 3-8-17, 227–228, 2007.
- [4] Matsuzaki *et al.*, Proc. Autumn Meet. Acoust. Soc. Jpn., 2-4-15, 471–474, 2007.
- [5] Matsuzaki *et al.*, J. Acoust. Soc. Jpn. (E), 17(3), 163–166, 1996.
- [6] Yoshikawa *et al.*, IEICE Trans. Fundamentals, vol.J81-A(3), 303–311, 1998.